PRESIDIO®

PROPOSAL RESPONSE

Massachusetts Technology Collaborative

AI Hub Data Commons Collaborative Consulting Services

Request for Proposals: 2026-JAH-01

Massachusetts Technology Collaborative
75 North Drive
Westborough, MA 01581-3340

Submit to: proposals@masstech.org

MASSACHUSETTS TECHNOLOGY COLLABORATIVE

www.presidio.com

PRESIDIO®

October 7, 2025

Massachusetts Technology Collaborative
Attn: Proposals
proposals@masstech.org

Subject: AI Hub Data Commons Collaborative Consulting Services
Request for Proposal (RFP) No. 2026-JAII-01

Thank you for the opportunity to respond to the Data Commons solicitation. The Massachusetts Technology Collaborative (MassTech) is seeking a partner to design and deploy the Data Commons, and Presidio is well-positioned to fulfill this role. We understand the scope of services outlined in the solicitation and are confident in our ability to deliver the desired solutions—exceeding expectations through both our products and services.

The Commonwealth widely regards Presidio as a trusted partner in the areas of Data Engineering, Analytics, and AI. We have successfully delivered modern data platform and AI solutions—leveraging AWS and Snowflake—for some of the Commonwealth's most significant initiatives. Our reputation is built on our ability to get things done, our flexible, collaborative approach, and our engineering excellence.

We believe our strong performance and proven track record with Commonwealth agencies highlight the consistent value we bring to every engagement.

As Presidio's designated Account Manager for the Commonwealth, I am the primary point of contact for any questions about our proposal. My direct contact information is:

Presidio Networked Solutions LLC
Dan Buckley, Senior Account Manager
Phone: 617-460-0959 | Email: dbuckely@presidio.com

If you have any questions or concerns regarding our proposal, please contact me via the phone number or email address provided above.

Sincerely,

Dan Buckley

Dan Buckley
Senior Account Manager

TABLE OF CONTENTS

1 Application Cover Sheet 6
2 Company Overview and Qualifications 8
2.1 Firm Overview 8
2.1.1 Presidio's Value Add 9
2.2 Qualifications and Experience 10
2.2.1 Experience with Synthetic Data and AI Fairness Tools 11
2.3 Use of Subcontractors 11
3 Technical Approach and Methodology 12
3.1 High-Level Architecture 12
3.2 Project Timeline 14
3.3 Project Management, Stakeholder Management & Coordination 14
3.3.1 Cost and Schedule Management 16
3.3.1.1 Cost Management 17
3.3.1.2 Schedule Management 17
3.3.1.3 Dependency Management 17
3.3.2 Deliverable Management 18
3.3.2.1 Presidio's Scrum Delivery 18
3.3.3 Quality Management 20
3.3.4 Risk and Issue Management 20
3.3.5 Communication Management 20
3.3.6 Testing 22
3.4 AI Engineering Lifecycle Management 23
3.4.1 Lifecycle Phases and Activities 23
3.4.2 Team Roles and Responsibilities 24
3.4.3 Benefits to MassTech 25
3.5 Solution Architecture 25
3.5.1 Guiding Principles and Constraints 27
3.5.2 Stakeholder Alignment 27
3.5.3 Business Architecture and Capabilities 27
3.5.4 Information System Architecture 29
3.5.4.1 Access Layer – DCC Massachusetts-AI Hub Web Application 29
3.5.4.2 Data Layer – Governed Data Platform 30
3.5.4.3 Data Hosting and Access Methods 30
3.5.4.4 Commons Credits System Integration 30
3.5.4.5 Security and Compliance 31

3.5.4.6 Massachusetts AI Data Platform 31
3.5.4.7 Synthetic Data Capability 33
3.5.5 Privacy, Security & Compliance 36
3.5.5.1 Cisco AI Defense Integration 37
3.5.5.2 Governance & Compliance Requirements 37
3.5.5.3 AI Fairness and Bias Management 37
3.5.5.4 Commons Credits System 39
3.5.5.5 Consumption Tiers 41
3.5.5.6 Security, Governance & Compliance 43
3.6 Proposed Alternatives and Enhancements 45
3.6.1 FinOps Intelligence: Forecasting, Automation, and Governance 45
3.6.2 Pilot User Feedback Loop 46
3.6.3 Strategic Industry Engagement 46
3.6.4 Presidio Ecosystem and Technology Alliances 46
3.6.5 Security and AI Guardrails 47
3.6.6 Architectural Principles: Open Source First, Cloud Native, and AWS Well-Architected 47
4 Key Personnel and Resumes 48
Resource 1: Solution Owner 49
Resource 2: Data Architect 51
Resource 3: Data Engineer 54
5 Costs 58
6 Past Performance 59
7 W-9 61
8 Authorized Application Signature Acceptance Form 64

TABLE OF EXHIBITS

Exhibit 1. Presidio's Project Management Approach 15
Exhibit 2. Presidio's Project Lifecycle Management Approach 16
Exhibit 3. Presidio's Quantitative Project Management Process 17
Exhibit 4. Presidio's Deliverable Management Process 18
Exhibit 5. Presidio's Scrum Delivery Approach 19
Exhibit 6. Presidio's Risk Management Process 20
Exhibit 7. Presidio's Communication Process 21
Exhibit 8. Presidio's Testing Approach 22
Exhibit 9. Presidio AI Engineering Lifecycle 24

Exhibit 10. Key Solution Architecture Capabilities 26
Exhibit 11. Core Business Capabilities 28
Exhibit 12. Stakeholder Roles and Responsibilities 28
Exhibit 13. Governance Alignment 29
Exhibit 14. Compliance Standards Matrix 44

1 APPLICATION COVER SHEET

Application Cover Sheet (Attachment A)

Response:

Presidio's completed Attachment A follows.

Attachment A

Application Cover Sheet

Name of Respondent Presidio Networked Solutions LLC			
Mailing Address 10 Sixth Road	City/Town Woburn	State MA	Zip Code 01801
Telephone 781-638-2373	Fax n/a	Web Address www.presidio.com	
Primary Contact for Clarification Dan Buckley		Primary Contact E-mail Address dbuckley@presidio.com	
Authorized Signatory James J. Doyle		Authorized Signatory E-mail Address jdoyle@presidio.com	
Legal Status/Jurisdiction(e.g., Massachusetts Corporation, LLC, LLP, etc.) a FL LLC		Respondents W-9 attached Respondents UEI No. MMTAF84Y4L79	

2 COMPANY OVERVIEW AND QUALIFICATIONS

A description of the firm responding to the RFP (including descriptions of proposed subcontractors, if any) and the firm's qualifications to perform the Services.

Response:

2.1 FIRM OVERVIEW

Presidio is one of the largest and most adept providers of advanced technology solutions and services. We are passionate about driving results for our clients, delivering the highest quality of products and services to help them unlock the unlimited potential of a completely connected world.

PRESIDIO IS...
A Trusted Advisor
Providing clarity to ensure the right technology investment choices
A Business Co-Creator
defining technology initiatives to execute organizational strategies
A Digital Disruptor
focused on the differentiated use of technology to support innovation
A Service Provider
with global resources to cover gaps in capability and accelerate successful outcomes
PRESIDIO

Presidio recognizes that in today's business climate, IT investments need to be strategic and practical to have immediate impact. Our holistic approach blends experience and expertise with proven processes and technologies to satisfy our clients' technical business objectives, providing significant value differentiation.

As a global solution provider, Presidio combines experience and stability with regional expertise and service. As your ally, we will be there to help you locally with the ability to cover all your business needs.

Presidio collaborates with technology vendors who drive innovation and are indisputable leaders of the IT industry. Our significant investment with select strategic partners allows us to achieve broad and deep sets of technical capabilities. Our 98 percent client retention rate and annual growth rates show our passion for driving client results.

We are more than a trusted partner and ally; we enable new thinking that benefits our clients now and in the future.

2.1.1 PRESIDIO'S VALUE ADD

PRESIDIO VALUE IS EVERYWHERE
We have a strong track record of solving difficult and ambitious challenges for our customers
Securely processing **every tax return filed in the US**
Broadcasting high-profile events like **Presidential Debates, Sporting Events**
Connecting **1 in 10 US college students** in 500+ institutions
Supporting life-saving systems in **500+ hospitals**
Synthesizing over **1 billion security events a month**

Presidio's superior expertise is our core strength. Our clients view Presidio as an extension of their organization proactively working with them to identify opportunities and apply solutions that improve technology challenges and satisfy business objectives. By taking the time to understand how our clients define success, Presidio designs enduring solutions that meet your immediate needs and prepares you for tomorrow.

One of Presidio's main competitive advantages is our ability to integrate complex, multi-technology environments. We have the technical expertise to navigate and assess unique IT challenges, interpret requirements, and transform the latest technology innovations into meaningful enhancements to our clients' existing infrastructure. We combine the strategic vision, planning skills, and business acumen typically found in only the largest consulting companies, with the actionable, results-oriented approach of systems integrators. We provide the right levels of support using the most cost-effective programs available.

2.2 QUALIFICATIONS AND EXPERIENCE

Presidio has extensive experience designing, building, and maintaining scalable, high-performance data architectures for government and enterprise clients. Our team specializes in data and analytics engineering, ETL pipelines, modern cloud-based data platforms, and metadata management.

We have worked with many government agencies to transform legacy data systems into cloud-native architectures, delivering enterprise-scale solutions that enhance data accessibility, interoperability, and data-driven decision-making and governance.

Presidio's expertise spans cloud-based data management, real-time analytics, AI-driven automation, and end-to-end BI implementations. We align technology with business goals to consistently deliver measurable results, pave the way for future innovation, and compete in rapidly evolving markets.

AREA OF EXPERTISE	DESCRIPTION
Business Intelligence (BI)	Deliver BI solutions enabling data-driven decision-making for clients across industries using enhanced executive dashboards and predictive analytics capabilities through leading BI tools such as Amazon QuickSight, Tableau and Microsoft Power BI.
Cloud Data Management	Execute large-scale data integrations and migrations to modern data platforms like AWS Redshift, Databricks, Microsoft Fabric, GCP BigQuery, and Snowflake. Develop robust data governance frameworks, master data management (MDM) strategies, and metadata repositories to improve data accuracy, consistency, and accessibility.
Analytics and Insights	Craft customized advanced analytics solutions that deliver actionable insights. Using custom-developed machine learning models for customer behavior analysis, time-series forecasting, and churn prediction. Provide solutions optimized for operational efficiency and enhanced customer engagement for clients. Build ML models for behavior prediction, forecasting, NLP, and real-time insights using frameworks like SageMaker, Vertex AI, and Databricks.
Automation	Spearhead automation initiatives that streamline workflows and reduce manual dependencies. Highlights include designing and implementing Robotic Process Automation (RPA) workflows and integrating API-driven automation across complex IT ecosystems.
Data Warehouse Solutions & Governance	Design and implement scalable, cloud-native data lakehouses using platforms like Amazon Redshift or Amazon SageMaker AI, Databricks, Google BigQuery, Microsoft Fabric, and Snowflake. Create centralized data repositories to support real-time analytics, simplify reporting, and drive informed decision-making at enterprise scale. Establish governance frameworks, including role-based access, data stewardship, cataloging, and audit trail management.

2.2.1 Experience with Synthetic Data and AI Fairness Tools

Presidio uses synthetic data in the ordinary course of business to test our solutions. This generally involves copying production data to lower environments and using RBAC to obfuscate data in sensitive fields. We propose using the open-source toolkit, AI Fairness 360, for detecting bias in generated synthetic data.

2.3 Use of Subcontractors

Mass Tech Collaborative's prior approval is required for any subcontracted services under any Agreement entered into as a result of this RFP. The selected Respondent will take all appropriate steps to assure that minority firms, women's business enterprises, and labor surplus area firms are used when possible. The selected Respondent is responsible for the satisfactory performance and adequate oversight of its subcontractors. Subcontractors are required to meet the same requirements and are held to the same reimbursable cost standards as the selected Respondent.

Response:

Presidio always attempts to use internal resources when delivering services to our customers; however, there may be occasions, due to niche technology needs or volume of services, when Presidio uses subcontractors to supplement our technology and professional services offerings to support customer projects.

When we use subcontractors, Presidio, as the prime contractor, manages their performance to ensure subcontractors deliver the same level of customer satisfaction as Presidio employees. We take full responsibility for scheduling, coordinating, and managing subcontractors, providing a seamless face and a single point of contact for the customer.

Our minority partner, PLEDG Global LLC (PLEDG Global), provides complementary and supplemental staffing, technology, and digital enterprise solutions for Presidio engagements. PLEDG Global is certified by the National Minority Supplier Development Council (NMSDC) as a Minority Business Enterprise (MBE).

3 TECHNICAL APPROACH AND METHODOLOGY

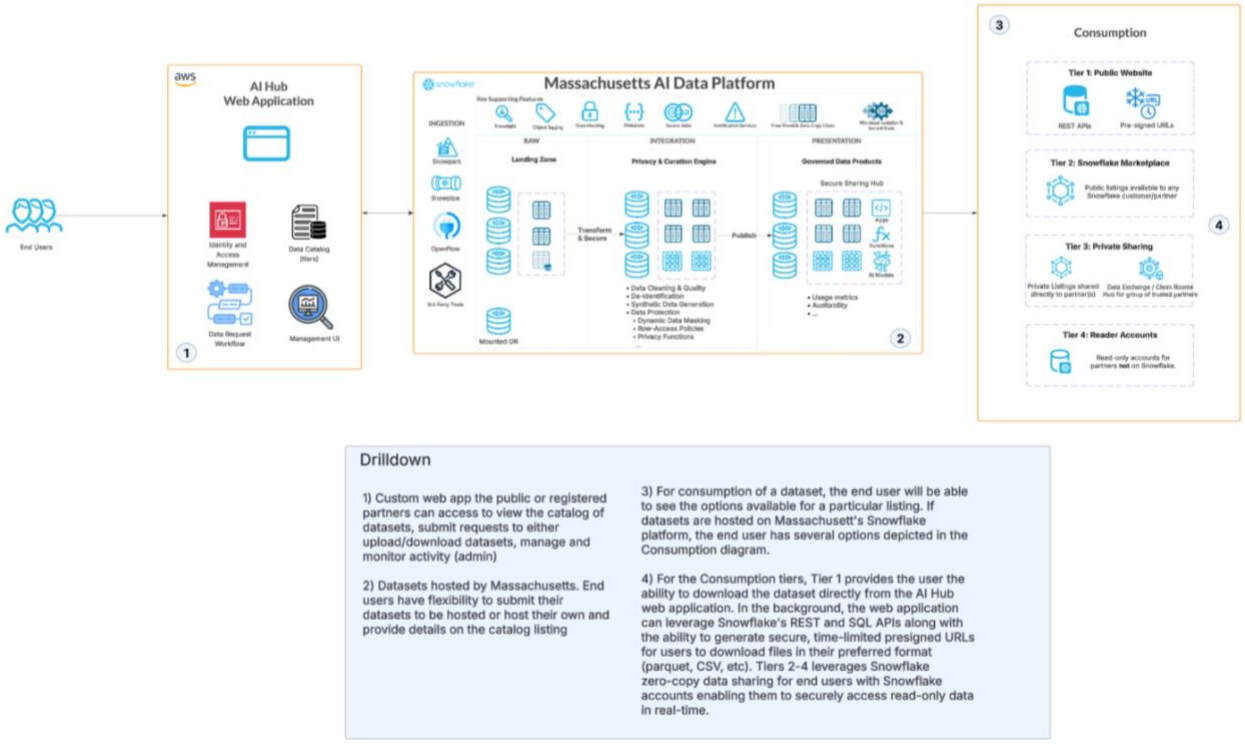
The proposed approach to providing the Services. Additionally, Respondents are invited to propose alternative(s) which provide substantially better or more cost-effective performance than achievable under the stated RFP scope of services.

Response:

Massachusetts Technology Collaborative (MassTech), on behalf of the Massachusetts AI Hub, seeks to engage a qualified partner for the design, development, and launch of the Data Commons Collaborative (DCC). This initiative is foundational to the Commonwealth's AI strategy and aims to centralize access to high-value datasets, enable synthetic data generation, and support ethical, fair, and secure AI development.

3.1 HIGH-LEVEL ARCHITECTURE

We propose a cloud/native architecture built on top of AWS and utilizing the Snowflake data platform. The key attributes of the architecture are shown in the diagram below:


AI Hub
Web Application
Massachusetts AI Data Platform
Consumption
Tier 1: Public Website
Tier 2: Snowflake Marketplace
Tier 3: Private Sharing
Tier 4: Reader Accounts
Drilldown
1) Custom web app the public or registered partners can access to view the catalog of datasets, submit requests to either upload/download datasets, manage and monitor activity (admin)
2) Datasets hosted by Massachusetts. End users have flexibility to submit their datasets to be hosted or host their own and provide details on the catalog listing
3) For consumption of a dataset, the end user will be able to see the options available for a particular listing. If datasets are hosted on Massachusett's Snowflake platform, the end user has several options depicted in the Consumption diagram.
4) For the Consumption tiers, Tier 1 provides the user the ability to download the dataset directly from the AI Hub web application. In the background, the web application can leverage Snowflake's REST and SQL APIs along with the ability to generate secure, time-limited presigned URLs for users to download files in their preferred format (parquet, CSV, etc). Tiers 2-4 leverages Snowflake zero-copy data sharing for end users with Snowflake accounts enabling them to securely access read-only data in real-time.

The following are key components of the proposed architecture:

COMPONENT	DESCRIPTION
AI Hub Web Application	The web UI layer will be implemented using the react/next.js framework and will be deployed on AWS. This web frontend will be main portal to enable for dataset discovery, submission, and management.

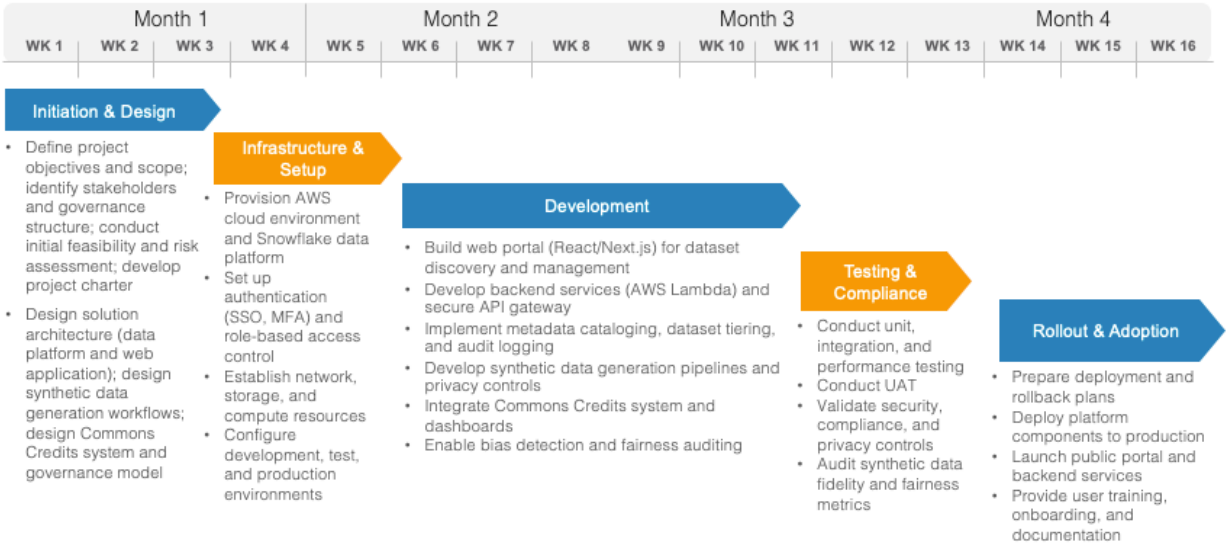
COMPONENT	DESCRIPTION
Snowflake Data Platform	The data layer will leverage Snowflake as the centralized system of record, supporting metadata cataloging, dataset tiering, synthetic data generation, and audit logging.
Synthetic Data Generation	Synthetic data will be generated using Snowflake's native synthetic data generation capabilities.
Bias Detection	Fairness auditing and bias mitigation will be handled via open-source frameworks including AIF360 and Fairlearn, with all assessments logged for transparency.
Common Credits System	The platform will incorporate a Commons Credits system for resource management and incentivization, and will be designed for extensibility, supporting future integration with federated data sources, blockchain audit trails, and advanced governance controls.

The architectural components will be deployed per the table below:

COMPONENT	HOSTED ON	ROLE/FUNCTION
AI Hub Web Application	AWS	User interface, dataset explorer
API Gateway	AWS	Secure API endpoint management
Backend Services	AWS Lambda	Business logic, Snowflake integration
Authentication	AWS Cognito	SSO, MFA, user management
Data Platform	Snowflake	Data storage, metadata, synthetic data, ledger
Synthetic Data Engine	Snowflake	Synthetic data generation, privacy enforcement
Bias Detection	Snowflake	AIF 360 Python package running within Snowflake's Snowsight notebook
Monitoring & Audit	AWS/Snowflake	Logging, compliance, dashboards

3.2 PROJECT TIMELINE

Below is a high-level timeline outlining the project's implementation. While the diagram presents the phases as progressing in a linear sequence for simplicity, not all tasks are iterative; many will occur in parallel throughout the project.


Month 1
Month 2
Month 3
Month 4
WK 1
WK 2
WK 3
WK 4
WK 5
WK 6
WK 7
WK 8
WK 9
WK 10
WK 11
WK 12
WK 13
WK 14
WK 15
WK 16
Initiation & Design
Define project objectives and scope; identify stakeholders and governance structure; conduct initial feasibility and risk assessment; develop project charter
Design solution architecture (data platform and web application); design synthetic data generation workflows; design Commons Credits system and governance model
Infrastructure & Setup
Provision AWS cloud environment and Snowflake data platform
Set up authentication (SSO, MFA) and role-based access control
Establish network, storage, and compute resources
Configure development, test, and production environments
Development
Build web portal (React/Next.js) for dataset discovery and management
Develop backend services (AWS Lambda) and secure API gateway
Implement metadata cataloging, dataset tiering, and audit logging
Develop synthetic data generation pipelines and privacy controls
Integrate Commons Credits system and dashboards
Enable bias detection and fairness auditing
Testing & Compliance
Conduct unit, integration, and performance testing
Conduct UAT
Validate security, compliance, and privacy controls
Audit synthetic data fidelity and fairness metrics
Rollout & Adoption
Prepare deployment and rollback plans
Deploy platform components to production
Launch public portal and backend services
Provide user training, onboarding, and documentation

3.3 PROJECT MANAGEMENT, STAKEHOLDER MANAGEMENT & COORDINATION

Presidio's project management approach incorporates standard processes and deliverables that have been implemented and refined on numerous Presidio projects. **Exhibit 1** depicts the core elements of our approach.

Exhibit 1. Presidio's Project Management Approach

The Presidio approach will enable us to complete the project in a cost-effective and timely manner and provide quality deliverables while greatly reducing risk. Our project management team arrives with established toolsets for:

- Personnel management
- Cost and schedule management
- Project planning
- Governance
- Data protection
- Deliverable management
- Quality management
- Risk management

Exhibit 2. summarizes part of Presidio's Program Life Cycle Management (PLCM) approach.

Exhibit 2. Presidio's Project Lifecycle Management Approach

PLCM APPROACH	BENEFIT TO MASSTECH
Our Solution Owner (SO) will work with MassTech to communicate views of the program performance utilizing status reports effectively and efficiently	Information for MassTech management and stakeholders on project progression
Cost and Schedule Management processes	Identify risks early and mitigate to deliver on-time and within budget
Deliverable and quality management processes built and refined from other Presidio projects and SOs	Consistent delivery of high-quality deliverables via standard templates and a rigorous review process
Effective and efficient program and communication processes provide synchronization between Presidio and MassTech.	Continuous and effective communications to provide project status and awareness to client stakeholders
Agile-Based Project Management	Presidio employs Agile-based project management methodologies to manage projects. This approach enables seamless shifts in project priorities between sprints, ensuring maximum flexibility.

3.3.1 COST AND SCHEDULE MANAGEMENT

Presidio uses our proven quantitative project management processes shown in **Exhibit 3** to:

- Manage cost and schedule to help the project management team predict outcomes to deliver on time and on budget.
- Estimate the effort and resources required to support changes to the project priorities and deliverables.
- Continuously measure and monitor planned versus actual schedule progress.
- Collect, analyze, review, and report project metrics at weekly and monthly intervals.

Presidio's business management teams generate monthly invoices and reports with data collected at regular monthly intervals. Presidio will investigate variances and work with MassTech to take the necessary corrective actions.

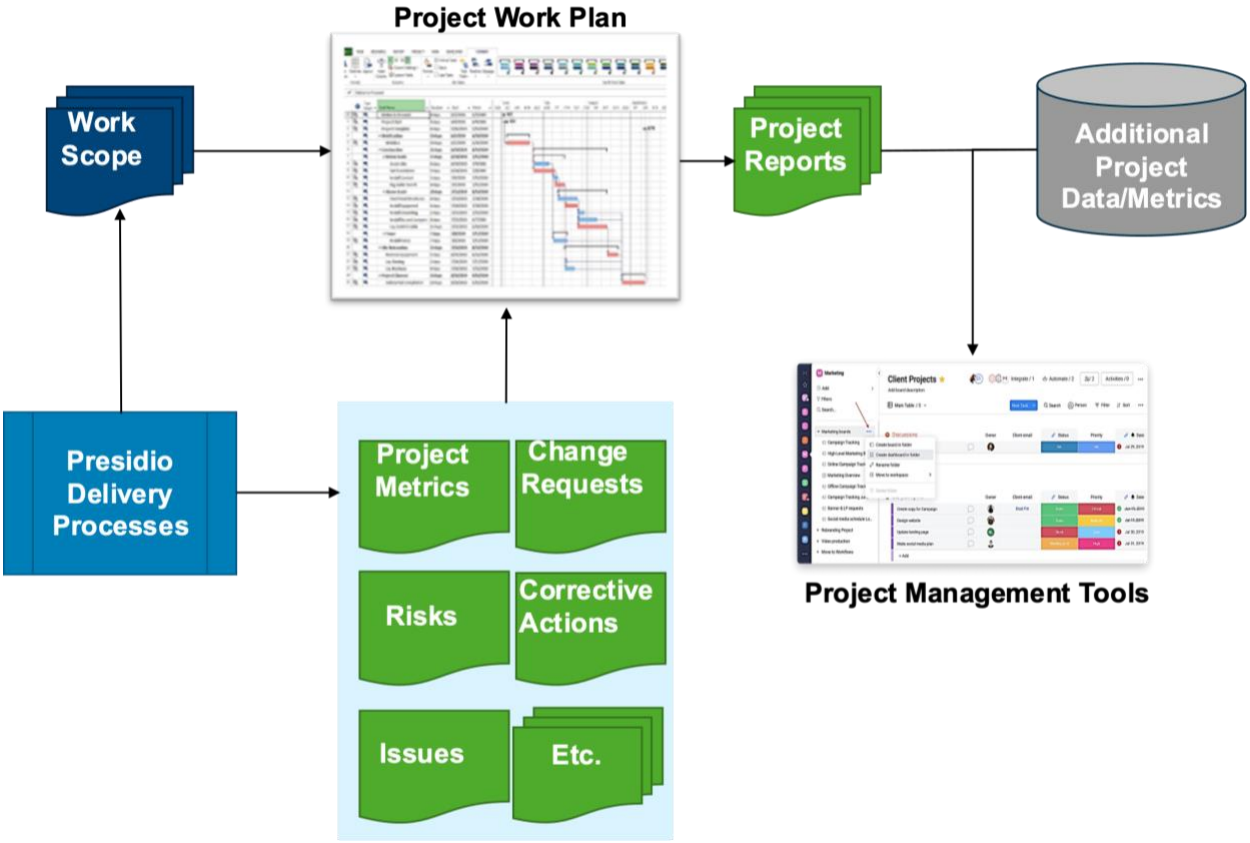


Exhibit 3. Presidio's Quantitative Project Management Process

3.3.1.1 Cost Management

The Presidio team tracks on-budget performance by monitoring measured planned versus actual costs. We collect, analyze, review and report financial metrics at regular intervals. Actual costs are monitored using time and expense information, provided through a standardized time reporting tool as we execute project work. The project management team will raise variances and recommend corrective actions to MassTech.

3.3.1.2 Schedule Management

Presidio will create, maintain, and regularly report status for the project schedule. Our team will work closely with all project stakeholders to create and manage the overall Project Schedule for the project. The Project Schedule review helps identify project dependencies and helps identify key project tasks, required artifact delivery and approval dates, and milestone dates.

3.3.1.3 Dependency Management

As in every project, proper dependency management is paramount in managing the schedule and risk for every project. We document inbound and outbound dependencies at project start and maintain their status in the project management log until the project's closure. If an identified

dependency is escalated to a risk/issue, that risk/issue will be documented in the corresponding log. Logged risks and issues will then be managed through their respective processes.

These and other Presidio management processes keep projects on schedule and within budget while ensuring quality deliverables.

3.3.2 DELIVERABLE MANAGEMENT

The Presidio SO is responsible for all project-management-related processes, including deliverable management. This process covers the research, creation, review, and final delivery of app Presidio project deliverables. **Exhibit 4** outlines the process that will govern creation status, review, and approval of Presidio project deliverables. There are three primary steps in this process. The initial step is to validate scope and acceptance criteria. The next step is to conduct working sessions with project stakeholders and develop a draft of the deliverables for initial review. The final step is the formal submission and review of the deliverables to MassTech.

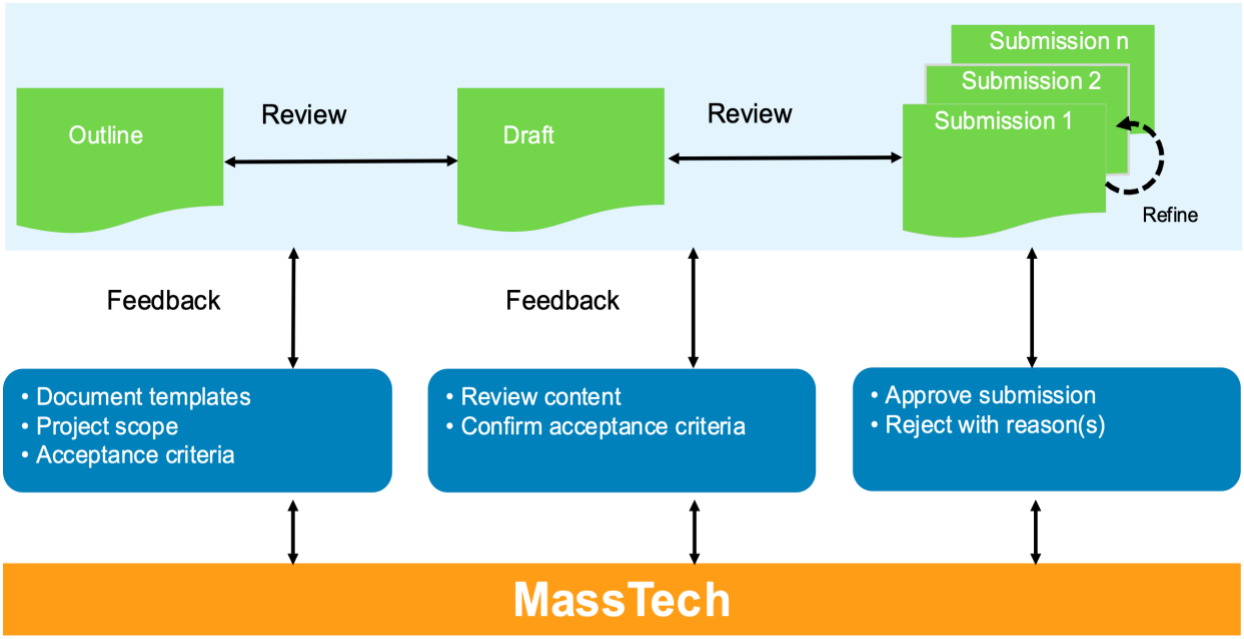


Exhibit 4. Presidio's Deliverable Management Process

3.3.2.1 Presidio's Scrum Delivery

Presidio is knowledgeable in Waterfall, Agile, and Hybrid delivery methodologies. We have successfully completed projects using these methods. depending on the project and the client's preference. The Presidio engagement team is well-versed in Agile and Scrum methodology and ceremonies. The Presidio project management team will be responsible for running the Agile ceremonies during the project. The Presidio project management team will adhere to the Agile delivery model, as shown in **Exhibit 5**, which includes sprint planning, daily stand-ups, backlog management and tracking, and sprint retrospectives. Presidio follows the Scrum Delivery model throughout each phase of the project.

During status meetings, we report project progress against plan, sprint completion metrics, backlog activity, and velocity to development teams, owners, and stakeholders.

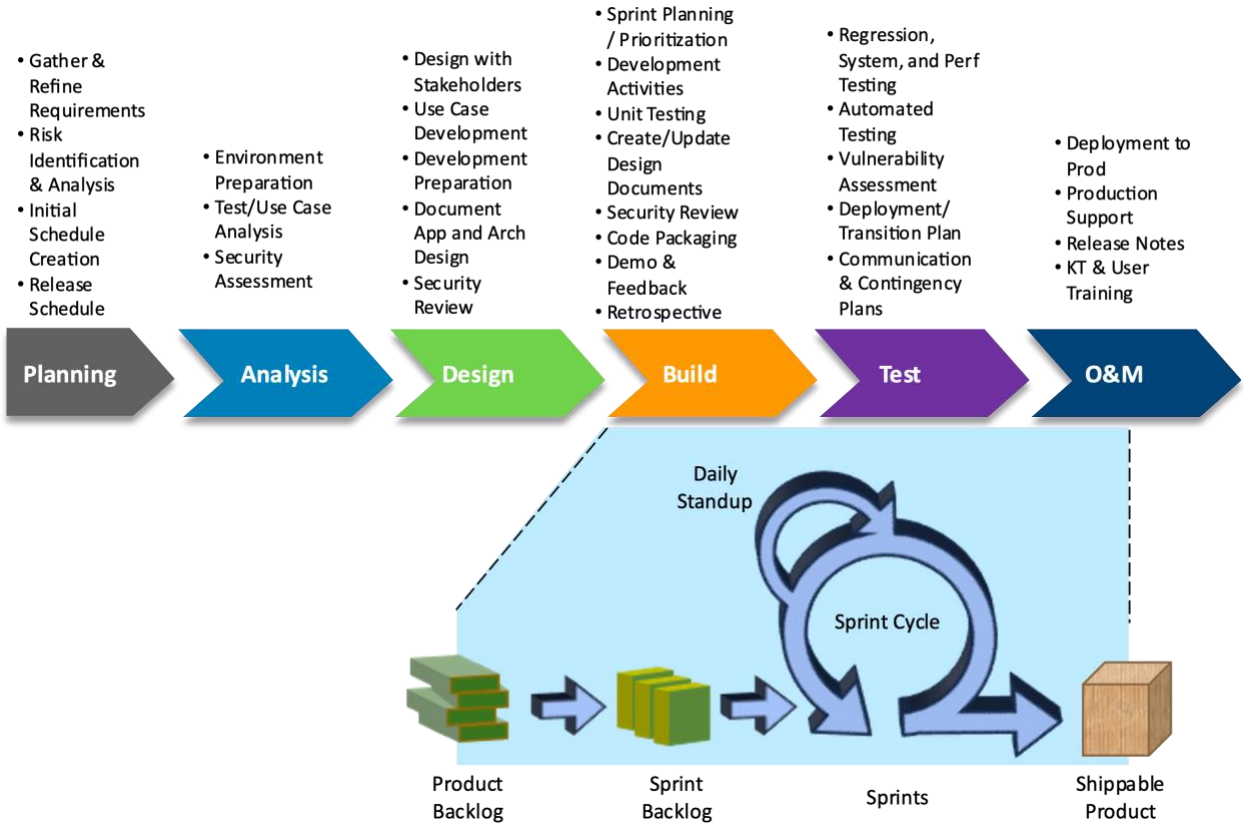


Exhibit 5. Presidio's Scrum Delivery Approach

Presidio's Agile method will provide an incremental and iterative approach to delivery. Our approach involves the formal Scrum Ceremonies that are listed in **Exhibit 5** above. Some of the Scrum Ceremonies we would implement during the engagement include:

- Daily Standups to review work completed, upcoming work for the day and notification of blockers for quick resolution.
- Cross-Team Standups and working sessions (as needed).
- Backlog refinement of prioritized Epics and Stories.
- Sprint Planning to define a commitment of time-boxed work in regular intervals.
- Internal and External Sprint Retrospectives for continuous feedback and improvement.
- Sprint demos aligned with project-status calls/reporting for ongoing feedback and knowledge transfer throughout the project.

3.3.3 QUALITY MANAGEMENT

The Presidio project management team will review the project deliverables and monitor processes to ensure high quality throughout the project life cycle. Presidio will also perform project quality reviews, as needed, on a regular basis. These quality reviews will be conducted by experienced resources who are not assigned to this project.

3.3.4 RISK AND ISSUE MANAGEMENT

Risk and Issue Management is key to a project's success. Presidio will implement a Risk and Issue Management process to track, address, mitigate, and monitor risks throughout the Project Lifecycle (PLC). We may customize our Risk Management and Issue Management processes to integrate with MassTech's existing risk and issue management process, if needed.

Exhibit 6 outlines our Risk and Issue Management Process. Our process begins with Risk Identification. In this stage, we evaluate previously identified risks and work with MassTech to identify new risks such as security, schedule, and operational risks to the project. After a risk is identified, we will perform an evaluation to determine its potential impact and probability of occurring. Based on the priority and ranking assignment for each risk within the Risk Evaluation stage, we will plan a mitigation approach for each risk. In the final stage we will track, monitor, and report risks. This process will recur on a regular basis throughout the PLC.

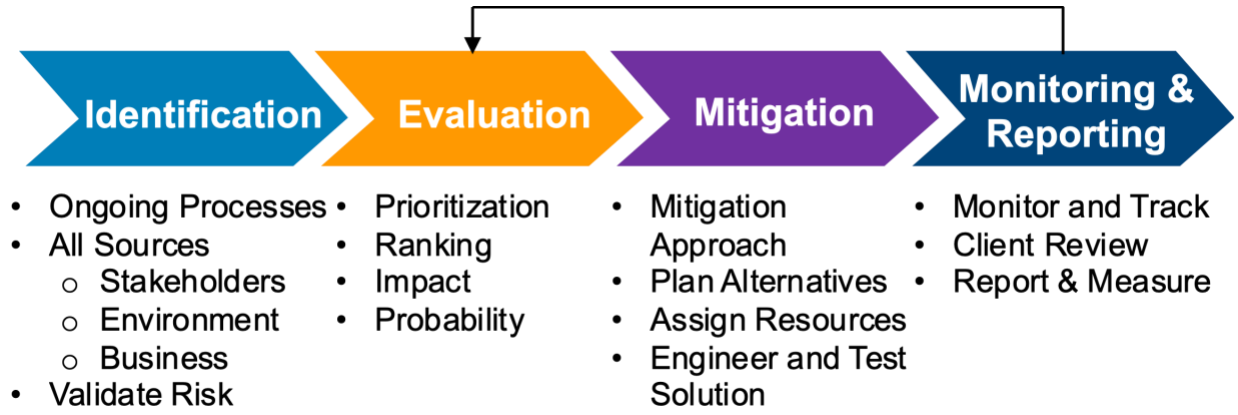


Exhibit 6. Presidio's Risk Management Process

3.3.5 COMMUNICATION MANAGEMENT

Communication is a cornerstone of successful project completion. Our Communication Plan will outline engagement with sponsors and stakeholders. The Presidio Communication Process will create and maintain a cohesive and collaborative project environment for both MassTech and Presidio team members assigned to the project.

Our communication plan includes:

- **Project Kickoff** - Introductions, review project expectations, scope, deliverables, ways of working and immediate next steps

- **Scrum Ceremonies** - Standups, cross-team standups and working sessions as needed, backlog refinement, sprint planning, internal and external sprint retrospectives, sprint demos (aligned with project status calls/reporting)
- **Governance/status calls with stakeholders**

Exhibit 7 illustrates how Presidio will promote continuous communication and awareness about the status and progress of the project.

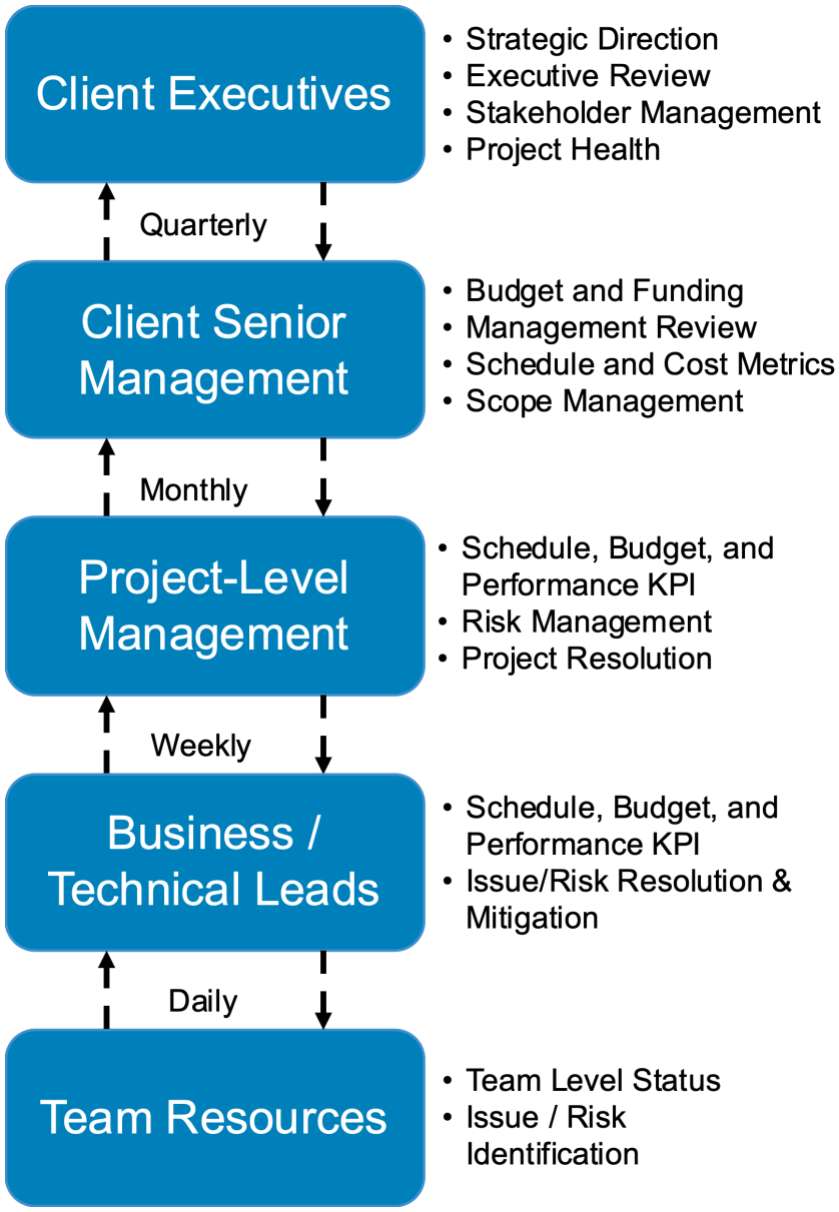


Exhibit 7. Presidio's Communication Process

3.3.6 TESTING

Our testing approach involves a series of tests, including functional testing, integration testing, performance testing, and user acceptance testing (UAT) **(Exhibit 8)**. Testing is a critical phase in the application development process since it ensures that the applications function correctly, meet performance benchmarks, and adhere to quality standards. This activity will require support from application business owners to conduct acceptance testing.

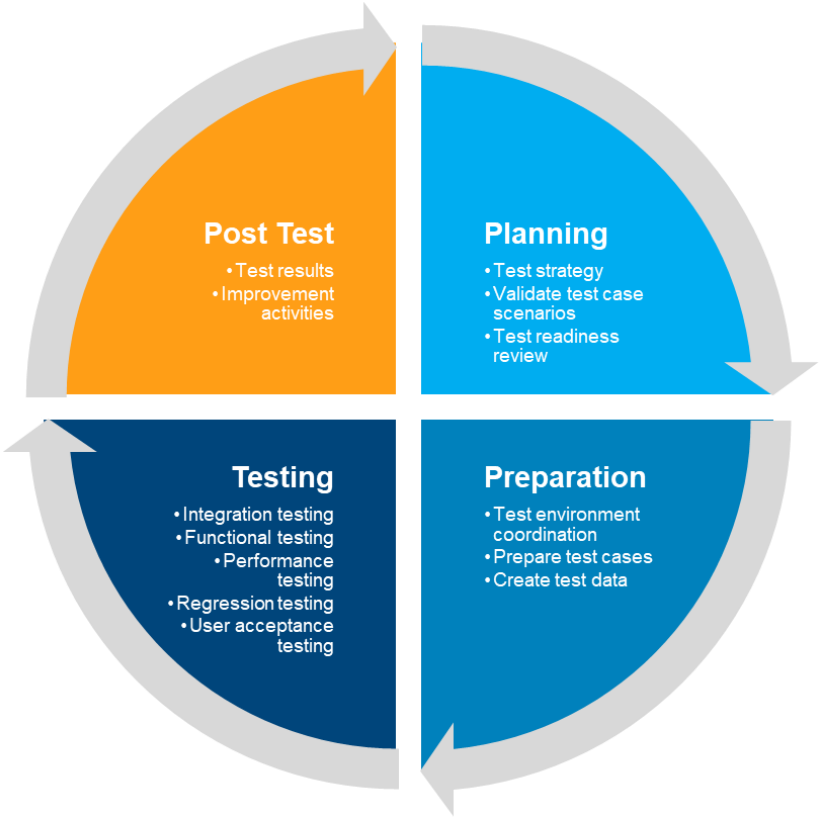


Exhibit 8. Presidio's Testing Approach

Unit Testing ensures individual components of a software application perform as expected. Unit tests are conducted by the development team and are designed to be independent to ensure that a specific section of the codebase, such as a function, method, or object, behaves as expected.

Integration Testing combines individual units, modules, or components of a software application and tests them as a group. The purpose is to expose faults in the interaction between integrated units to ensure that they work together correctly.

Functional Testing evaluates the software against its functional requirements/specifications. It involves testing each function of the application by providing appropriate input and verifying the output, ensuring the software performs its intended functions correctly.

Performance Testing evaluates the performance and scalability of a system or application under various conditions. It aims to identify bottlenecks, measure responsiveness and stability, and ensure the system can handle the expected workload, including the number of users or transactions.

Regression Testing is the process of re-running functional and non-functional tests to ensure that previously developed and tested software still performs as expected after a change.

User Acceptance Testing is the opportunity for MassTech's subject matter experts to ensure the solution meets the functional and non-functional requirements.

3.4 AI Engineering Lifecycle Management

Presidio's Responsible AI Engineering Lifecycle is designed to ensure that all AI-related activities within the DCC initiative are conducted ethically, securely, and in alignment with Massachusetts' governance and compliance standards. This lifecycle integrates seamlessly with our broader project management methodology and supports the delivery of trustworthy, inclusive, and sustainable AI systems.

3.4.1 Lifecycle Phases and Activities

PHASE	ACTIVITIES AND GOVERNANCE
Assess	• Evaluate current AI readiness across data, workforce, and infrastructure. • Identify strategic business priorities and ethical considerations. • Conduct gap analysis for data governance, privacy, and fairness.
Adopt	• Design and implement AI/Analytics solutions aligned with DCC goals. • Define governance policies, SOPs, and responsible usage guidelines. • Enable hybrid deployment models and secure integration workflows.
Practice	• Operationalize AI solutions with continuous monitoring and updates, • Implement fairness audits, bias mitigation workflows, and explainability tools. • Maintain audit trails, consent management, and compliance dashboards.
Live	• Transition AI systems to production with full governance enforcement. • Support stakeholder onboarding, training, and feedback loops. • Enable continuous improvement through participatory audits and community engagement.

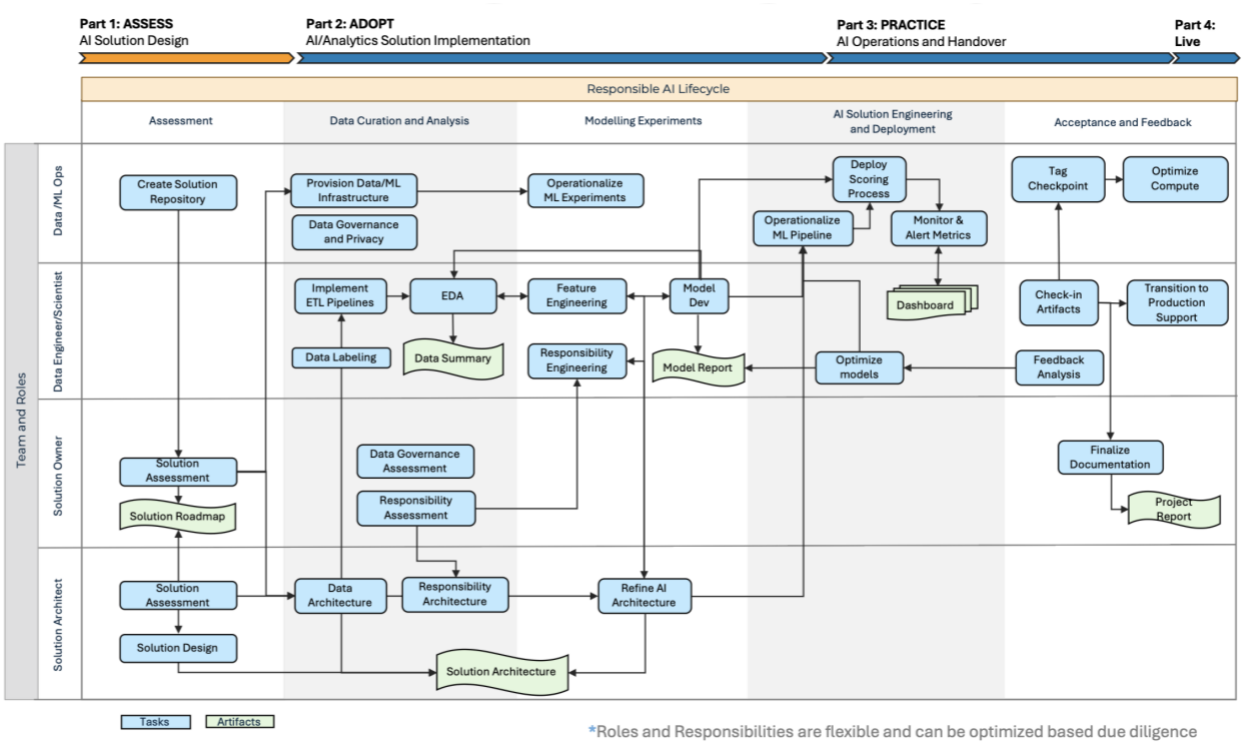


Exhibit 9. Presidio AI Engineering Lifecycle

3.4.2 TEAM ROLES AND RESPONSIBILITIES

PHASE	ENTITY	ROLES	RESPONSIBILITIES
Assess	Presidio	• Solution Owner • AI Architect	• Solution Design and Road-mapping • Technology Recommendations
	MassTech	• Business Architect • Enterprise Architect • Stakeholder	• Strategic and Business Requirements • Knowledge share on infrastructure and dependencies
Adopt	Presidio	• Data Engineer • AI/ML Ops • BI	• Data Engineering and Curation • Analytics and AI Experimentation
	MassTech	• SME • Data Architect • Business Analyst • Security Architect • Governance Lead	• Subject Matter Expertise on Data and Business Processes • Security, Governance, and Privacy Requirements
Practice	Presidio	• Data Scientists, ML Ops	• Model Management

PHASE	ENTITY	ROLES	RESPONSIBILITIES
	MassTech	• Operations Manager • Service Delivery Manager	• Operational Support
Live	Presidio	• BI • Platform Owner	• Transition to production and governance handoff
	MassTech	• Application Architect • Risk Manager	• Ongoing support and feedback

3.4.3 BENEFITS TO MASSTECH

Presidio's delivery-centric approach to governance ensures measurable value to MassTech through:

- **Accelerated AI Enablement**: Centralized access to curated datasets and synthetic data capabilities supports rapid experimentation and ethical AI development.
- **Operational Efficiency**: Agile-based delivery, structured governance, and the Commons Credits system promote scalable, cost-effective platform usage.
- **Security and Compliance**: Embedded controls (RBAC, MFA, encryption, audit logging) ensure alignment with HIPAA, GDPR, 201 CMR 17.00 (MA), and NIST AI RMF.
- **Responsible AI Practices**: Engineering teams leverage managed tools and frameworks for fairness, explainability, and governance, including:
  - **AIF360**, **Fairlearn** – Bias detection and mitigation
  - **SHAP**, **LIME** – Model explainability
  - **Datasheets for Datasets**, **Open Ethos** – Documentation and ethical transparency
  - **Cisco AI Defense –** Responsible AI Operations with SIEM integrated inline

This integrated lifecycle supports MassTech's strategic goals for trustworthy, inclusive, and future-ready AI infrastructure.

3.5 SOLUTION ARCHITECTURE

The proposed architecture directly supports MassTech's goals for the Data Commons Collaborative (DCC), including centralized access to high-value datasets, ethical AI enablement, secure data sharing, and synthetic data experimentation. It is designed to meet the functional and non-functional requirements while remaining extensible for future capabilities such as model tracking, sustainability metrics, and federated access.

The proposed architecture supports MassTech's core objectives for the DCC:

- **Centralized Access** to curated, high-value datasets from Massachusetts agencies, global repositories, and synthetic data generators.
- **Ethical AI Enablement** through integrated fairness auditing, bias mitigation, and governance workflows.
- **Secure and Governed Data Sharing** via role-based access control (RBAC), multi-factor authentication (MFA), and audit trails.
- **Synthetic Data Capability** for privacy-preserving experimentation and model development.
- **Commons Credits System** to incentivize contributions and manage platform usage.

Our design is modular, extensible, and compliant with FAIR principles, Massachusetts cybersecurity policies, and accessibility standards (WCAG 2.1 AA). It integrates metadata management, RBAC, and audit capabilities, while supporting federated data discovery and synthetic data pipelines.

Exhibit 10. Key Solution Architecture Capabilities

CAPABILITY	DESCRIPTION
Metadata and Cataloging	Implementation of a searchable, standards-compliant metadata system (e.g., DCAT, Schema.org) with support for dataset versioning, provenance tagging, and automated enrichment.
Synthetic Data Generation	Integration of tools such as SDV, Synthea, and CARLA, with support for privacy-preserving synthesis, validation workflows, and audit traceability.
AI Fairness Framework	Deployment of bias detection and mitigation tools (e.g., AIF360, Fairlearn), with dashboards for fairness reporting and compliance tracking.
Commons Credits System	Design of a centralized ledger to manage incentive-based usage, with support for role-based quotas, transaction logging, and future extensibility.
Security and Governance	Enforcement of RBAC, MFA, encryption, and ethical data handling policies, aligned with HIPAA, GDPR, and Massachusetts 201 CMR 17.00 (MA).
User Experience and Accessibility	Delivery of a responsive, multilingual portal with dataset explorer, sandbox environments, and training modules for diverse user groups.

The solution is designed to interoperate with MGHPCC's AICR infrastructure and supports future integration with Snowflake and other state-managed platforms.

This proposal reflects our commitment to delivering a scalable, secure, and inclusive data commons platform that accelerates AI innovation across Massachusetts.

The architecture separates the user-facing Access Layer from the governed Data Layer, enabling flexible consumption models while maintaining centralized control. The Access Layer is built on AWS and handles identity, permissions, and UI. The Data Layer is powered by Snowflake, serving as the system of record and governance engine, with support for zero-copy sharing, synthetic data integration, and multimodal formats.

LAYER	CONTROLS
Access (MA-AI Hub Web Application on AWS)	• Role-based access, SSO, and consent workflows • Dataset browsing, metadata search, and tiered access • WCAG 2.1 AA compliance and multilingual support
Data (Snowflake Data Cloud)	• Centralized governance and policy enforcement • Secure hosting of curated datasets • Integration with synthetic data tools and Snowflake-native sharing

3.5.1 Guiding Principles and Constraints

The architecture is guided by the following principles:

PRINCIPLE	DESCRIPTION
Modularity	All components are loosely coupled and independently deployable.
Governance by Design	Security, auditability, and ethical use are embedded at every layer.
Open Standards First	Preference for DCAT, Schema.org, FAIR principles, and open-source tooling (e.g., CKAN, Dataverse).
Hybrid Hosting	Supports MGHPCC/AICR and cloud-native services (e.g., Snowflake, AWS).
Extensibility	Designed to accommodate future capabilities such as blockchain, federated modeling, and model usage tracking.

Constraints include:

- Data residency within U.S. regions, with East Coast preference.
- No remote access to Commonwealth data from outside the continental U.S.
- Accessibility compliance (WCAG 2.1 AA) and multilingual readiness.

3.5.2 Stakeholder Alignment

This architecture framework ensures alignment with:

- MassTech and MA AI Hub governance and procurement policies.
- State agency data stewards and their operational constraints.
- Flexible access and sandbox environments for academic and industry partners.
- Transparent, ethical, and inclusive data access for public users.

3.5.3 Business Architecture and Capabilities

This section outlines the core business capabilities, stakeholder roles, and governance alignment necessary for successful implementation and adoption.

Exhibit 11. Core Business Capabilities

CAPABILITY	DESCRIPTION
Dataset Curation & Stewardship	Vetting, tagging, and maintaining high-value datasets for inclusion in the DCC
Metadata Management	Ensuring completeness, provenance, and FAIR compliance across all datasets
Synthetic Data Enablement	Supporting privacy-preserving experimentation and model development
Commons Credits Administration	Managing issuance, redemption, and audit of platform usage incentives
Access Governance & Consent	Enforcing tiered access control using Public, Registered, and Restricted dataset classifications (DCC Section 8.4 and 13.1).
AI Fairness Oversight	Auditing datasets and models for bias, equity, and transparency
Stakeholder Engagement	Coordinating with agencies, researchers, and industry for onboarding and feedback
Training & Capacity Building	Delivering onboarding, workshops, and certification for ethical AI practices

Exhibit 12. Stakeholder Roles and Responsibilities

ROLE	RESPONSIBILITIES
Administrators	Oversee platform governance, access control, and policy enforcement
Data Stewards	Curate, redact, and validate datasets; ensure metadata quality
Researchers	Discover, access, and analyze datasets; contribute metadata and feedback
Industry Partners	Use data for innovation; contribute datasets and participate in pilot programs
Educators & Students	Leverage datasets and sandbox environments for learning and curriculum development
Public Users	Browse and access public datasets; provide feedback and suggestions
Challenge Participants	Granted time-limited access for innovation sprints and competitions
Compliance Auditors	Independently assess compliance, privacy, and ethical safeguards
Fairness Auditors	Evaluate datasets and models for bias; document fairness metrics and impact
Consent Managers	Oversee platform governance, access control, and policy enforcement, including management of tokenized, revocable, and time-bound consent workflows for restricted datasets (DCC Requirements Section 12.0).
Governance Committee	Approve policies, oversee Commons Credits, and ensure ethical platform operation

Exhibit 13. Governance Alignment

GOVERNANCE ELEMENT	DESCRIPTION
Hybrid Governance Model	Centralized coordination via a Data Access & Use Committee, with distributed approval frameworks for partner-held data
Ethical Oversight	Embedded workflows for bias impact documentation, consent management, and fairness auditing
Policy Compliance	Adherence to HIPAA, GDPR, 201 CMR 17.00 (MA), and emerging AI governance frameworks (e.g., NIST AI RMF) Includes accessibility compliance with WCAG 2.1 AA and multilingual readiness as required in DCC Sections 5.9 and 6.2.
Stakeholder Inclusion	Representation from state agencies, academia, industry, and community groups in governance processes

3.5.4 INFORMATION SYSTEM ARCHITECTURE

The Information Systems Architecture for the Massachusetts AI Hub – Data Commons Collaborative (DCC) is designed to support secure, scalable, and equitable access to curated and federated datasets. It separates the Access Layer (user-facing portal) from the Data Layer (governed data platform), enabling centralized governance, metadata management, synthetic data generation, and tiered access control.

This architecture directly supports the DCC's goals as outlined in Sections 6.0–13.0 of the DCC Requirements document, including FAIR compliance, privacy-preserving collaboration, and extensibility for future AI capabilities.

3.5.4.1 Access Layer – DCC Massachusetts-AI Hub Web Application

CAPABILITY	DESCRIPTION	DCC REQUIREMENTS
Role-Based Access Control (RBAC)	Assigns dataset access based on user roles (e.g., researcher, public viewer, admin)	Section 13.1
Consent Management	Tokenized, revocable, time-bound consent workflows for restricted datasets	Section 12.0
Dataset Explorer	Search, filter, and preview datasets with multilingual support and WCAG 2.1 AA compliance	Sections 5.2, 6.2
API Key Issuance	Secure issuance of keys for programmatic access to hosted datasets	Section 6.1
Dynamic Access Options	Dataset detail pages display access methods based on user role and dataset tier	Section 6.2

3.5.4.2 Data Layer – Governed Data Platform

CAPABILITY	DESCRIPTION	DCC REQUIREMENTS
Metadata Management	Descriptive, technical, and provenance metadata using DCAT, Schema.org, Dublin Core	Section 8.2
Dataset Tiering	Public, Registered, and Restricted access levels with audit trails	Sections 8.4, 13.1
Synthetic Data Generation	Integrated tools (e.g., SDV, Synthea, CARLA) with privacy scoring and lineage tracking	Section 10.0
FAIR Compliance	Supports Findable, Accessible, Interoperable, Reusable data principles	Section 8.1
Federated Access	Metadata-only discovery and secure access to external datasets	Section 12.0
Audit Logging	≥3 years retention of access events, synthetic generation, and metadata edits	Section 13.1

3.5.4.3 Data Hosting and Access Methods

METHOD	DESCRIPTION	TARGET USERS	DCC REQUIREMENTS
RESTful APIs	Programmatic access to machine-readable formats (CSV, JSON, Parquet)	Developers, researchers	Section 6.1
Pre-Signed URLs	Secure, time-limited links for large file downloads	Analysts, researchers	Section 6.1
Reader Accounts	Free Snowflake access for partners without licenses	Strategic partners	Section 12.0
Data Clean Rooms	Privacy-preserving collaboration without raw data exposure	Healthcare, sensitive domains	Section 10.2
Snowflake Marketplace	Public and private listings for live data sharing	Snowflake ecosystem	Section 6.1

3.5.4.4 Commons Credits System Integration

CAPABILITY	DESCRIPTION	DCC REQUIREMENTS
Ledger Architecture	Centralized, auditable transaction log with metadata tagging	Section 7.1
Usage Events	Tracks downloads, API calls, synthetic data generation, metadata enrichment	Section 7.2–7.3
Access Control	Role-based quotas, expiration policies, throttling rules	Section 7.6

CAPABILITY	DESCRIPTION	DCC REQUIREMENTS
Governance Tools	Admin dashboards for credit issuance, usage reporting, and policy enforcement	Section 7.5
Future-Ready Design	Pluggable backend for blockchain integration and sustainability metrics	Section 7.7

3.5.4.5 Security and Compliance

CAPABILITY	DESCRIPTION	DCC REQUIREMENTS
Authentication	SSO, MFA, RBAC with dataset-level permissions	Section 13.1
Encryption	Data encrypted at rest and in transit	Section 13.2
Logging & Auditing	Access logs retained ≥3 years; SIEM integration supported	Section 13.1
Compliance Standards	HIPAA, GDPR, 201 CMR 17.00 (MA), NIST AI RMF	Section 12.3
Incident Response	24-hour breach notification, audit trail support	Section 13.2

3.5.4.6 Massachusetts AI Data Platform

The Massachusetts AI Data Platform will serve as the secure, governed back end for the DCC, enabling centralized data hosting, metadata management, synthetic data generation, and tiered access control. Built on Snowflake and integrated with open-source cataloging tools, the platform is designed to support FAIR principles, privacy compliance, and scalable collaboration across public, academic, and commercial stakeholders.

1. **Metadata Cataloging and Discovery**: The platform will implement a searchable, standards-compliant metadata system using **DCAT**, **Schema.org**, and **Dublin Core**, with support for:
    - Descriptive, technical, and provenance metadata
    - Version control and lineage tracking
    - Faceted and full-text search
    - Integration with persistent identifiers (e.g., DOIs via DataCite)
2. **Dataset Tiering and Access Governance**: Datasets will be classified into **Public**, **Registered**, and **Restricted** tiers, with role-based permissions and audit trails. Access workflows will be enforced via Snowflake's native RBAC and masking policies.
3. **Synthetic Data Integration**: The platform will support privacy-preserving synthetic data generation using tools such as **SDV**, **Synthea**, and **CARLA**, with:
    - Secure ingestion pipelines
    - Privacy scoring and de-identification

- Metadata tagging and provenance tracking
- Scalable compute environments for model training

4. **Commons Credits Ledger**: A centralized, auditable ledger will track credit issuance, usage, and redemption events. The platform will log:
   - Dataset uploads, API calls, synthetic data requests
   - Metadata enrichment and labeling contributions
   - Model training and compute usage
5. **Governance and Policy Enforcement**: The platform will enforce compliance with **HIPAA**, **GDPR**, **201 CMR 17.00 (MA)**, and emerging AI governance frameworks (e.g., **NIST AI RMF**). It will support:
   - Consent workflows
   - Audit logging (≥3 years retention)
   - Encryption at rest and in transit
   - Incident response and breach notification
6. **Interoperability and Extensibility**: The platform will support integration with external systems via RESTful APIs and SDKs, enabling:
   - Federated access to external datasets
   - Linked data and FAIR compliance
   - Future blockchain integration for audit trails
7. **Dataset Curation and Integration**: Presidio will support curation of initial datasets, ensuring metadata completeness and adherence to FAIR data principles.
   - Our team will develop automated ingestion pipelines, version control, and provenance tagging to support ongoing data integration and traceability.
8. **Dataset Integration Matrix**: This Dataset Integration Matrix has been developed to support the Massachusetts AI Hub – Data Commons Collaborative (DCC) initiative. It synthesizes guidance from:
9. **DCC Requirements Document:**
   - **Section 2.0** – Services Required: Dataset curation, ingestion pipelines, metadata completeness, and FAIR principles.
   - **Section 5.0** – Key Functional Requirements: External dataset integration, synthetic data capabilities, and metadata standards.
   - Appendix A: Illustrative list of potential datasets for onboarding.
10. **DCC RFP Q&A Document:**

- **Questions 93–94, 197, 206, 265:** Clarify expectations around dataset sourcing, volume, update frequency, and vendor responsibilities**.**
- **Questions 153, 154, 221, 223:** Metadata enrichment, lineage, and user experience expectations.
- **Appendix A:** Lists potential datasets for onboarding**.**
- **Appendix B:** Includes relevant Q&A responses that inform integration priorities, metadata standards, and ingestion methods.

This matrix will guide the technical implementation of dataset onboarding, metadata enrichment, and interoperability across the DCC platform. It also includes update frequency and download links to support planning and automation of ingestion workflows.

ATTRIBUTE	MASSGIS LAND USE (2016)	MA PUBLIC HEALTH DATA	MBTA TRANSIT LOGS	MA EDUCATION OUTCOMES	MASSDEP CLIMATE DATA
Source Entity	MassGIS	MA Dept. of Public Health	MBTA	MA Dept. of Education (DESE)	MassDEP
Domain	Geospatial	Healthcare	Transportation	Education	Environmental
Format & Integration	Shapefile, GeoTIFF/ Batch via API	HL7 FHIR, CSV/ Secure batch ingest	CSV, GeoJSON, KML/ Streaming (Phase 2)	Excel, CSV, PowerBI / Batch	AERMET (*.sfc, *.pfl), CSV/ Batch
Metadata Standard	DCAT, Schema.org	Schema.org	DCAT	Schema.org	Dublin Core
Access Tier	Public	Restricted	Internal	Restricted	Public
Synthetic Support & Lineage	Yes (CARLA)/ Dataset-level	Yes (Synthea)/ Column-level planned	Planned/ Dataset-level	Yes/ Dataset-level	Yes/ Dataset-level
Download Link	https://www.mass.gov/info-details/massgis-data-2016-land-coverland-use	https://www.mass.gov/dph-data-library	https://mbta-massdot.opendata.arcgis.com/	https://www.mass.gov/info-details/dese-directory-of-datasets-and-reports	https://www.mass.gov/info-details/meteorological-data

3.5.4.7 Synthetic Data Capability

To support privacy-preserving experimentation and public access to high-utility datasets, the DCC platform will integrate a Synthetic Data Capability that generates statistically representative datasets from sensitive seed data. This capability will be deployed within a secure, isolated environment and governed by strict privacy, compliance, and audit controls. Synthetic datasets will be made available via the DCC portal and APIs, with no exposure of original source data.

The platform will leverage Snowflake's native generate_synthetic_data procedure and integrate external tools such as SDV, Synthea, and CARLA for domain-specific synthesis. All synthetic data generation will be tracked via the Commons Credits ledger and governed by AI Guardrails and OWASP AI Security standards.

Key Capabilities

1. **Public Access to Synthetic Data Only**: The portal will expose only fully synthetic datasets. No original seed data will be accessible via the UI or APIs.
2. **Metadata Transparency**: Each synthetic dataset will include metadata on generation method, synthesis date, and statistical fidelity metrics.
3. **Statistical Fidelity Enforcement**: Synthetic datasets will replicate key statistical properties (e.g., means, variances, correlations) of seed data. Automated utility checks (e.g., the Kolmogorov–Smirnov goodness-of-fit test) will ensure >90% fidelity.
4. **Search & Filter Features**: Users can search synthetic datasets by domain, topic, generation date, and statistical metrics.
5. **Differential Privacy Support**: Synthetic generation will optionally use DP-GAN or DP-VAE with configurable budgets.
6. **Seed Data Destruction**: After generation, seed data will be masked and purged. Only the trained model and synthetic output will be retained.
7. **Reproducibility & Traceability**: Each generation event will log the random seed, model config, and DP parameters for audit and reproducibility.
8. **Standard AI Security and Guardrails**: Presidio recommends infrastructure level, standards-based AI security and guardrails model to ensure robust security, privacy, governance, compliance, risk management and mitigation, legislative regulations, operations, incident management, SIEM and audit requirements.

MA-AI Hub Features and Extensions

To support the Synthetic Data Capability architecture, our solution includes the following UI components within the MA-AI Hub.

FEATURE / EXTENSION	UI COMPONENT	DESCRIPTION
Synthetic Data Request Panel	Tool Selection	• Users can choose from supported synthetic data generators (e.g., SDV, Synthea, CARLA).
	Configuration Options	• Privacy level • Fidelity threshold • Target variables and join keys
	Credit Cost Display	• Shows estimated Commons Credits required for the selected configuration.

FEATURE / EXTENSION	UI COMPONENT	DESCRIPTION
	Security & Rate Controls	• Backend-specific guardrails for tool execution • Rate limiting and quota enforcement • Role-based access enforcement
Synthetic Data Preview & Metadata	Dataset Tags	• Clearly indicate synthetic origin and generation tool used.
	Provenance Display	• Show lineage from seed data, model configuration, and synthesis parameters.
	Statistical Summary	• Display key metrics (e.g., mean, variance, correlation) and fidelity score.
	Differential Privacy Indicators	• Visual cues for datasets generated with DP mechanisms.
Synthetic Dataset Explorer	Search & Filter Options	• Domain, topic, generation date • Statistical fidelity metrics • Privacy level
	Dataset Cards	• Metadata summary • Download or API access link (if permitted) • Usage policy tags (e.g., restricted for high-risk AI)
Governance Dashboard (Admin Only)	Audit Logs	• View synthesis history, model versions, and DP parameters.
	Bias Reports	• Access fairness metrics and mitigation actions.
	Security Alerts	• Notifications for model drift, data drift, or re-identification risk.
	Tool Execution Monitoring	• Track backend performance and quota usage.
Common Credits Integration	Credit Ledger View	• Display credit usage per synthetic generation event.
	Quota Enforcement	• Alert users when nearing limits or expiration.
	Request Workflow	• Enable approval process for restricted synthetic generation tasks.
Security Dashboard Integration	Guardrail Enforcement Panel	• Admins can configure and monitor AI Guardrails (Security, Safety, Privacy) applied to synthetic data workflows.
	Threat Detection Alerts	• Real-time alerts for prompt injection, model inversion, and unauthorized access to synthetic datasets.
	Audit Trail Viewer	• Visualize lineage, model configurations, DP parameters, and red teaming results for each synthetic dataset.

FEATURE / EXTENSION	UI COMPONENT	DESCRIPTION
Policy Management Interface	Policy Usage Enforcement	• Define and enforce acceptable use policies for synthetic data, blocking high-risk AI applications (e.g., social scoring).
	RBAC	Assign granular permissions for synthetic data generation, access, and governance tasks.
	Shadow AI Visibility Panel	Detect and report unauthorized synthetic data tools or queries executed against Snowflake.

Governance

FEATURE	DESCRIPTION
RBAC & MFA Enforcement	Access to the SDG engine will be restricted to essential roles using MFA and short-lived credentials.
Audit Logging & Version Control	All generation events, model versions, and hyperparameters will be logged and retained for ≥3 years.
Bias Mitigation	Fairness metrics (e.g., Disparate Impact Ratio) and algorithmic interventions (e.g., re-sampling) will be applied to reduce bias in synthetic data.
Usage Policy Enforcement	The portal will block synthetic data access for prohibited use cases (e.g., social scoring, biometric surveillance).
Model Protection	Generative models will be treated as sensitive IP, protected via access controls and watermarking.

Integration Architecture

COMPONENT	DESCRIPTION
Snowflake SDG Engine	Synthetic data will be generated using Snowflake's GENERATE_SYNTHETIC_DATA procedure with JSON-based configuration.
Secure Execution Environment	The SDG engine will run in a network-isolated VPC, separate from the public portal layer.
External Tool Integration	SDV, Synthea, CARLA, and other tools will be containerized and orchestrated via Snowflake stages or external compute environments.
Seed Data Validation	Anomaly detection will be applied to seed data to prevent poisoning or manipulation.

3.5.5 Privacy, Security & Compliance

FEATURE	DESCRIPTION
Regulatory Alignment	Synthetic data workflows will comply with HIPAA, GDPR, 201 CMR 17.00 (MA), and the EU AI Act.

FEATURE	DESCRIPTION
Re-identification Risk Assessment	Final synthetic datasets will be scanned for similarity to source data using MIA and model inversion simulations.
Consent Enforcement	Synthetic datasets derived from restricted seed data will inherit consent tokens and expiration policies.
Data Minimization	Sensitive variables will be masked or pseudonymized before training. Seed data will be purged post-generation.

3.5.5.1 Cisco AI Defense Integration

Architecture-Level AI Security and Guardrails

SECURITY FEATURE/GUARDRAIL	DESCRIPTION
Federated Security Monitoring	Cisco AI Defense integrates with Snowflake via Splunk Data Fabric, enabling real-time monitoring of synthetic data generation, access, and lineage.
AI Threat Intelligence Overlay	Cisco Talos and MITRE ATLAS-aligned threat feeds are applied to Snowflake-hosted synthetic datasets to detect adversarial patterns, prompt injection risks, and model inversion attempts.
Algorithmic Red Teaming	Synthetic data models hosted in Snowflake can be automatically stress-tested using Cisco's AI red teaming engine to identify vulnerabilities.
Runtime Guardrails	Cisco AI Defense enforces dynamic security, privacy, and safety guardrails during synthetic data generation and access workflows

3.5.5.2 Governance & Compliance Requirements

REQUIREMENT	DESCRIPTION
Shadow AI Detection	Cisco AI Defense identifies unsanctioned synthetic data tools or unauthorized Snowflake queries, ensuring only approved pipelines are used.
Model Validation & Drift Detection	Continuous validation of synthetic models for performance degradation, fairness drift, and compliance violations.
EU AI Act & OWASP Alignment	Cisco AI Defense supports external audit readiness and OWASP AI Guardrails enforcement for Snowflake-hosted synthetic data.

3.5.5.3 AI Fairness and Bias Management

The AI Fairness and Bias Management capability ensures that datasets and models developed or hosted within the DCC platform are equitable, transparent, and aligned with ethical standards. This capability supports bias detection, mitigation, and continuous monitoring across synthetic and real datasets, enabling responsible AI development in high-impact domains such as healthcare, education, and public policy.

The platform will integrate open-source fairness tools (e.g., AIF360, Fairlearn), support algorithmic audits, and enforce governance workflows for ethical review and community engagement. All fairness assessments will be logged, versioned, and made accessible via dashboards to promote transparency and accountability.

Key Capabilities

CAPABILITY	DESCRIPTION
Bias Detection & Auditing	Automated and manual audits of datasets and models to detect disparities across protected attributes (e.g., race, gender, age).
Fairness Evaluation Tools	Integration of fairness metrics (e.g., Equal Opportunity, Disparate Impact Ratio) into model development workflows.
Bias Mitigation Techniques	Support for pre-processing (e.g., reweighting), in-processing (e.g., fairness-aware learning), and post-processing (e.g., outcome adjustment).
Ethical Oversight Workflow	Formal review process for datasets and models, including bias impact statements and reviewer annotations.
Community Engagement	Participatory audits, feedback channels, and documentation of social context to ensure inclusive outcomes.
Education & Capacity Building	Training modules, workshops, and best practice repositories to build a community of responsible AI practitioners.

Governance

FEATURE	DESCRIPTION
Auditability & Transparency	All fairness assessments and mitigation actions will be logged, versioned, and made publicly accessible via dashboards.
Role-Based Access Control (RBAC)	Fairness auditors, reviewers, and data stewards will be assigned roles with scoped permissions.
Policy Compliance	Alignment with HIPAA, GDPR, Algorithmic Accountability Act, and EU AI Act.
Fairness Oversight Board	High-risk applications will be reviewed by a designated board with authority to approve or reject datasets/models.

Integration Architecture

COMPONENT	DESCRIPTION
Tool Integration	Support for AIF360, Fairlearn, Themis-ML, SHAP, LIME, and Jurity for fairness metrics, explainability, and mitigation.
Development Environment Compatibility	Integration with Jupyter, MLFlow, TensorBoard, and Airflow for fairness monitoring in model pipelines.
Monitoring & Drift Detection	Use of Evidently AI and WhyLogs to detect fairness drift and performance degradation over time.

COMPONENT	DESCRIPTION
Explainability & Documentation	Use of Datasheets for Datasets and Open Ethos for documenting provenance, ethics, and stakeholder engagement.

Privacy & Compliance

FEATURE	DESCRIPTION
Protected Attribute Handling	Sensitive attributes used in fairness audits will be access-controlled and privacy-preserving.
Bias Audit Privacy	Fairness evaluations will not expose individual-level data or increase re-identification risk.
Consent Alignment	Fairness assessments will respect consent tokens and opt-out policies associated with datasets.
Legal Safeguards	All fairness workflows will comply with anti-discrimination lAWS and emerging AI regulations.

UI Requirements

To support this capability, the following UI features will be implemented in the MA-AI Hub:

UI FEATURE	FUNCTIONALITY
Fairness Audit Panel	• Select dataset or model for audit • Choose fairness metrics and protected attributes • View audit results and disparity scores • Submit bias impact statement
Bias Mitigation Workflow	• Apply mitigation techniques (pre-, in-, post-processing) • Compare fairness vs. performance trade-offs • Document mitigation rationale
Fairness Dashboard	• View audit history and versioned reports • Monitor fairness drift across projects • Access reviewer annotations and approvals
Community Feedback Interface	• Submit annotations or social context notes • Participate in participatory audits • Flag datasets/models for ethical review
Training & Education Portal	• Access fairness tutorials and workshops • Browse case studies and best practices • Track certification progress for responsible AI

3.5.5.4 Commons Credits System

The Commons Credits System is a foundational mechanism within the Data Commons Collaborative (DCC) that incentivizes contributions, governs equitable access, and promotes

sustainable usage of shared AI infrastructure. The system is anchored by a centralized, auditable ledger and is designed to support a wide range of contributors and consumers across the Massachusetts AI ecosystem.

Credit Flow Architecture

The Commons Credits System operates through three core stages: **Earn**, **Ledger**, and **Spend**.

Earn Credits

Credits can be earned by a diverse set of ecosystem participants, including:

- Industry partners
- Academic researchers
- Startups
- Grassroots organizations
- Educators and students
- Individual contributors

Credits are issued through three primary channels:

- **Contributions**: Uploading datasets, labeling data, enriching metadata, or submitting tools and documentation.
- **Grants**: Allocated by the MA AI Hub to support research, education, or public interest initiatives.
- **Purchase**: Optional mechanism to acquire credits using real currency (future phase).

Each earning event is logged in the Commons Credits Central Ledger with metadata tags (e.g., data_contribution, metadata_enrichment, challenge_participation).

Commons Credits Central Ledger

At the heart of the system is a secure, centralized ledger that:

- Tracks issuance, redemption, and transfer of credits
- Integrates with identity and access management (SSO, API tokens)
- Maintains a historical transaction log with timestamps, user roles, and purpose tags
- Supports future extensibility for blockchain-based audit trails

The ledger is governed by administrative policies and is accessible via dashboards for transparency and reporting.

Spend Credits

Credits can be redeemed for access to premium platform capabilities, including:

Compute Resources:

- **General Model Use**: Training or inference on shared infrastructure
- **Energy-Efficient Model Use**: Incentivized lower-cost access for sustainable AI workloads

Storage Resources:

- **Tiered Storage Access**: Scales with dataset size and retention duration

Additional use cases include:

- Access to restricted or high-value datasets
- Synthetic data generation
- Unlocking third-party applications or services

The system enforces **quotas, expiration policies, and throttling rules** to ensure fair and efficient usage.

Governance and Transparency:

Administrative users can:

- Set credit issuance policies and user/org quotas
- Monitor usage patterns and generate reports
- Adjust thresholds for compute/storage consumption
- Review and approve grant-based allocations

All governance actions are logged and auditable, supporting compliance with DCC Section 7.5.

User Experience Integration:

The MA-AI Hub Web Portal will include:

- A **Commons Credits Dashboard** for users to view balances, transaction history, and eligible actions
- Guidance on how to earn credits through contributions
- Alerts for quota thresholds and expiration
- Multilingual and accessible UI (WCAG 2.1 AA)

3.5.5.5 Consumption Tiers

The DCC platform supports a tiered consumption model that aligns resource access with user roles, dataset sensitivity, and Commons Credits availability. This model ensures equitable access, incentivizes efficient usage, and enables sustainable scaling of compute and storage infrastructure.

TIER	DESCRIPTION	TARGET USERS	COMMONS CREDITS REQUIRED
I – Public	Open datasets freely accessible via portal or API.	General public, educators, students	None
II – Registered	Requires user authentication (SSO) and acceptance of terms of use.	Researchers, developers, civic tech users	Low or none
III – Restricted	Access gated by approval workflows, consent tokens, or credit thresholds.	Government agencies, healthcare orgs	Medium to high
IV – Premium	High-value datasets, synthetic data tools, or compute-intensive services.	Challenge participants, advanced users	High

Consumption Categories

CATEGORY	FEATURE	DESCRIPTION
Data Access	Direct Download:	CSV, JSON, Parquet formats
	API Access	RESTful endpoints with rate limits
	Zero-Copy Sharing	Snowflake-native access for partners
	Clean Rooms:	Privacy-preserving analysis environments
Compute Services	Model Training & Inference	General-purpose compute Energy-efficient compute (discounted credits)
	Synthetic Data Generation	Tool-specific quotas and rate controls Credit-based gating for high-fidelity or DP-enabled synthesis
Storage	Short-Term	Temporary storage for active projects
	Long-Term	Archival storage with credit-based scaling
	Versioned Datasets	Credits scale with number of versions retained

Enforcement & Governance

SOLUTION ELEMENT	DESCRIPTION
Quota Management	Role-based quotas enforced at the API and UI layers
Throttling Rules	Prevent abuse of compute/storage resources
Expiration Policies	Credits expire after defined periods unless renewed
Audit Logging	All consumption events logged for ≥3 years

User Experience Integration

SOLUTION ELEMENT	DESCRIPTION
Dynamic Access Display	Dataset detail pages show available access methods based on user role and credit balance
Commons Credits Dashboard	Users can track consumption by category (e.g., compute, storage, API)
Alerts & Notifications	Users are notified when nearing quota or expiration thresholds

3.5.5.6 Security, Governance & Compliance

The DCC platform is designed with security, governance, and compliance embedded by design, ensuring that all data access, processing, and sharing activities align with Massachusetts state policies, federal regulations, and emerging AI governance frameworks.

Security Architecture

CAPABILITY	DESCRIPTION
Authentication	Single Sign-On (SSO) with Multi-Factor Authentication (MFA) for sensitive data
RBAC	Role-Based Access Control for datasets, tools, and Commons Credits
Encryption	AES-256 encryption at rest and TLS 1.2+ in transit
Logging & Auditing	≥3 years retention of access logs, synthetic generation, and credit usage
Intrusion Detection	Integrated with SIEM for real-time threat monitoring
Incident Response	24-hour breach notification and audit trail support

Governance Framework

ELEMENT	DESCRIPTION
Hybrid Governance Model	Centralized Data Access & Use Committee + distributed partner approvals
Consent Management	Tokenized, revocable, time-bound consent workflows for restricted datasets
Policy Enforcement	Dataset tiering (Public, Registered, Restricted), usage terms, and audit logs
Commons Credits Oversight	Admin dashboards for quota setting, issuance policies, and usage reporting
Ethical Oversight	Bias impact statements, fairness audits, and community engagement workflows

Compliance Standards

The platform is compliant with:

- **HIPAA** (Health data privacy)

- **GDPR** (EU data protection)
- **201 CMR 17.00 (MA)** (Massachusetts data security regulations)
- **NIST AI RMF** (AI risk management framework)
- **EU AI Act** (for future alignment)

Exhibit 14. Compliance Standards Matrix

CATEGORY	STANDARD/FRAME WORK	PURPOSE	APPLICABLE TO
Data Privacy	HIPAA	Protects health information and governs PHI handling	Health datasets, synthetic data generation
	GDPR	Ensures data subject rights, consent, and data minimization	EU-origin data, consent workflows
	201 CMR 17.00 (MA)	Massachusetts regulation for safeguarding personal information	All user and dataset access
	FERPA (as applicable)	Protects student education records	Educational datasets
AI Governance	NIST AI RMF	Risk-based framework for trustworthy AI	Model development, fairness auditing
	Algorithmic Accountability Act (U.S.)	Promotes transparency and fairness in automated systems	Bias mitigation, audit trails
	EU AI Act (anticipated alignment)	Regulates high-risk AI systems and bias controls	Synthetic data, model training
Accessibility	WCAG 2.1 AA	Ensures digital accessibility for users with disabilities	MA-AI Hub Web Portal
	Multilingual Readiness	Supports inclusive access across diverse user groups	Portal UI, dataset explorer
Security & Auditability	Encryption (at rest & in transit)	Protects data confidentiality and integrity	All hosted and synthetic datasets
	RBAC + MFA	Enforces secure, role-based access	Dataset access, synthetic tools, admin dashboards
	Audit Logging (≥3 years)	Enables compliance reviews and forensic analysis	Access events, credit usage, synthetic generation
	Incident Response (24-hr notification)	Ensures timely breach reporting and mitigation	Platform-wide
Ethical Use & Consent	Tokenized, Revocable Consent	Manages access to restricted datasets with traceability	Consent workflows, synthetic data derivation

CATEGORY	STANDARD/FRAME WORK	PURPOSE	APPLICABLE TO
	Usage Policy Enforcement	Prevents misuse (e.g., social scoring, surveillance)	Dataset access, synthetic data tools
	Bias & Fairness Audits	Ensures equitable outcomes and transparency	High-impact datasets and models

AI-Specific Guardrails

GUARDRAIL	DESCRIPTION
Bias Mitigation	Integrated fairness tools (AIF360, Fairlearn) with audit trails
Synthetic Data Controls	Differential privacy, lineage tracking, and usage policy enforcement
Shadow AI Detection	Cisco AI Defense integration to detect unauthorized tools or data flows
Red Teaming & Drift Detection	Continuous validation of models and synthetic datasets

User Experience & Transparency

FEATURE	DESCRIPTION
Consent Dashboards	Users can view, manage, and revoke dataset access
Audit Trail Viewer	Admins can trace data lineage, access events, and credit transactions
Policy Documentation	Publicly accessible terms of use, data handling policies, and fairness guidelines

3.6 Proposed Alternatives and Enhancements

To further strengthen the DCC Massachusetts AI Hub's operational maturity, Presidio proposes an integrated enhancement package that combines intelligent financial governance, stakeholder engagement, and architectural resilience. These enhancements are designed to ensure the platform remains cost-effective, scalable, and aligned with evolving standards in ethical AI and data commons infrastructure.

3.6.1 FinOps Intelligence: Forecasting, Automation, and Governance

Presidio recommends extending the existing cloud financial optimization strategy with AI-driven forecasting and automated FinOps workflows. This capability ensures that financial operations are actively managed in real time, reducing risk and improving accountability.

RECOMMENDATION	DESCRIPTION/BENEFIT
Predictive Budgeting	DeepSight AI models forecast cloud spend with high accuracy, enabling proactive budget planning and scenario modeling.
Anomaly Response Automation	Cost spikes trigger automated remediation actions (e.g., rightsizing, shutdowns), governed by policy-based controls.

RECOMMENDATION	DESCRIPTION/BENEFIT
Governance Integration	All actions are logged, auditable, and aligned with DCC's compliance and ethical use policies.

3.6.2 PILOT USER FEEDBACK LOOP

To ensure the platform meets real-world needs, a structured pilot phase will be conducted with representative users from participating institutions. This loop ensures the platform evolves in lockstep with user expectations and operational realities.

- Early access to core features.
- Feedback via surveys, interviews, and usage analytics.
- Iterative refinement cycles based on pilot insights.
- A feedback dashboard to track issue resolution and feature evolution.

3.6.3 STRATEGIC INDUSTRY ENGAGEMENT

To ensure future readiness and maintain alignment with national and global policies and standards, the project will:

- Host quarterly roundtables with academic and industry stakeholders.
- Participate in AI governance and data commons forums.
- Establish advisory relationships with leaders in data interoperability, privacy, and ethical AI.

3.6.4 PRESIDIO ECOSYSTEM AND TECHNOLOGY ALLIANCES

Presidio will bring its strategic partner ecosystem to bear, enabling secure, scalable, and compliant infrastructure:

- Managed services for observability, compliance, and zero-trust architecture.
- Hybrid cloud deployment models tailored to institutional needs.
- Joint innovation labs focused on data federation and AI governance.

Key technology partners include:

- **Cisco**: Network security, observability, and AI Defense.
- **Snowflake**: Secure data sharing and cross-cloud analytics.
- **Splunk**: Real-time monitoring and operational intelligence.
- **AWS**: Cloud-native infrastructure and AI/ML services.

3.6.5 Security and AI Guardrails

To ensure responsible AI use and data protection:

- Role-based access controls and continuous monitoring.
- Differential privacy and data minimization enforcement.
- AI guardrails for model transparency, bias detection, and auditability.
- Alignment with NIST AI RMF and other federal guidelines.
- Cisco AI Defense integration for threat detection, anomaly response, and adversarial protection.

3.6.6 Architectural Principles: Open Source First, Cloud Native, and AWS Well-Architected

The solution architecture will adhere to modern engineering principles:

- **Open Source First**: Preference for community-supported platforms to ensure transparency and extensibility.
- **Cloud Native**: Microservices-based, containerized deployments for scalability and resilience.
- **AWS Well-Architected Framework**: Alignment with the five pillars—Operational Excellence, Security, Reliability, Performance Efficiency, and Cost Optimization.

4 KEY PERSONNEL AND RESUMES

Resumes or bios of key personnel that will be performing the Services.

Response:

For the sake of continuity, consistency, and quality assurance, Presidio's key project personnel are assigned and committed to each client project throughout its duration. Since project timing and specific personnel availability are undetermined at the time of proposal submission, Presidio assigns project teams upon contract award. At that time, specific personnel will be designated by name and assigned to these positions, ensuring that each brings the necessary expertise to meet the needs of the project.

This approach allows us to maintain flexibility while ensuring that the best-suited professionals are selected to drive the success of our partnership.

On the following pages, we provide resumes that exemplify the qualifications and experience of three key Presidio resources:

- Solution Owner
- Data Architect
- Data Engineer

RESOURCE 1: SOLUTION OWNER

RESOURCE SUMMARY

Results-driven technical leader with 15+ years of experience in complex, high-stakes environments across technology, engineering, delivery, and product innovation. Proven ability to build and scale solution ownership functions, drive strategic initiatives, and deliver impactful results in fast-paced settings. Skilled in AI/ML strategy, chaos engineering, technical program management, and cross-functional collaboration. Known for innovation leadership—holder of two patents—and recognized with awards for excellence in communication, crisis management, and team leadership.

CERTIFICATIONS

- Networking Foundations: Networking Basics
- Cloud Architecture: Core Concepts
- Cert Prep: Scrum Master
- Scrum: The Basics
- Enterprise Architecture Foundations

CURRENT ROLE

Presidio, Engineering Director

06/2025 – Present

- Developing the Solution Ownership practice to drive end-to-end delivery alignment across engineering, product, and business stakeholders.
- Responsible for enabling cross-functional teams to deliver scalable, AI-enabled solutions by defining technical strategy, managing execution, and ensuring business impact.
- Owning solution delivery from inception to value realization- Bridging business needs with technical execution across the SDLC
- Partnering with product, engineering, and client leadership to ensure alignment
- Coaching solution owners and delivery leads to drive clarity, accountability, and measurable outcomes
- Driving adoption of modern delivery practices including AI-augmented development, DevOps, and cloud transformation
- Leading strategic initiatives around enterprise modernization and innovation enablement

ENGAGEMENT HIGHLIGHTS

Solution Owner – AI/ML Strategy Initiative, Insurance Industry

Led the strategic development, approach, and implementation of an AI/ML strategy resulting in the successful integration of a claims adjudication Intelligence Platform into existing infrastructure. Project evolved into US Corporate AI strategy with the development of a Responsible AI framework focused on building trust.

Engagement Lead – Chaos Engineering, Financial Industry

Led development of innovative Chaos Engineering competency within a highly regulated environment, implementing resiliency automation and infrastructure to support the design/execution of experiments and testing as Scrum master and Engagement Lead for a team of twenty. Additional support provided for education, onboarding, and executive leadership.

Founder - Delivery Leadership Cloud Guild

Building off of the Boston Cloud Community, founder of the delivery leadership chapter with the vision of growing our roster of technically competent Team Members motivated by the pursuit of innovative cloud technology with a focus on bringing best in class solutions to our clients.

ADDITIONAL RELEVANT EXPERIENCE

Slalom, Senior Deliver Principal – Tech & Innovation

06/2022-06/2025

EDUCATION

Northeastern University

Master of Business Administration (MBA),

High Technology and Innovation (MS) (2013 - 2015)

University of Massachusetts Dartmouth

BSME, Mechanical Engineering (2007 - 2011)

RESOURCE 2: DATA ARCHITECT

- Innovative and detail-oriented Data Architect with experience designing and implementing robust data solutions for Fortune 500 companies.
- Proficient in data modeling, database design, and big data technologies, with a strong emphasis on optimizing data workflows and ensuring data integrity.
- Adept at translating business requirements into technical solutions, leveraging expertise in cloud platforms SQL/NoSQL databases, ETL processes, and data warehousing.
- Demonstrated success in leading cross-functional teams, managing complex projects, and driving strategic data initiatives to enhance business decision-making.
- Committed to staying ahead of industry trends and continuously improving data architecture practices.

CERTIFICATIONS

- SnowPro Core Certification
- Microsoft certified Azure Fundamentals

AREAS OF EXPERTISE

- Apache Parquet
- Apache Cassandra
- Apache Kafka
- Apache Flink
- AWS Rekognition
- Apache Hudi
- AWS EMR
- AWS SageMaker
- AWS AppFlow
- Tableau
- AWS RDS
- AWS Redshift
- AWS Glue
- AWS Data Pipeline
- AWS CloudFormation
- AWS Storage Gateway
- AWS EventBridge
- AWS Database Migration Service
- Apache Sqoop
- Apache Hive
- Apache Oozie
- Apache Hadoop
- Apache Spark
- AWS Aurora
- AWS Kinesis
- CPython
- PySpark
- AWS Step Functions
- Python
- AWS DynamoDB
- AWS Lake Formation
- AWS Athena
- AWS Lambda
- HashiCorp Terraform
- AWS Key Management Service (KMS)
- AWS DataSync

SKILLS OVERVIEW

- Strong business and communication skills as well as knowledge of the big data industry
- Systems development lifecycle and project management
- Creative and analytical problem-solving

- Database and cloud computing design, architecture, and data lakes
- ML, predictive modeling, and NLP
- Data visualization, data migration, and data modeling

CURRENT ROLE

Presidio, Data Architect

01/2021 – Present

ENGAGEMENT HIGHLIGHTS

08/2022-Present
Data Architect, Financial and Insurance
Rehost On-Prem SQL Server Cluster to Cloud

- Work with database architect to understand current ecosystem.
- Create dependency diagram to be utilized for migration
- Evaluate other Microsoft services like SSIS and SSRS
- Create architecture diagram for rehosting SQL server

01/2021 to 06/2022
Data Architect, Automotive Aftermarket
Enterprise Data Platform

- Create an inventory of the data necessary to build and implement a Finance analytics dashboard.
- Envision data pipelines and how data will flow through the enterprise.
- Evaluate current data management technologies and what additional tools are needed.
- Determine upgrades and improvements to current data architectures.

04/2020 to 01/2021
Senior Data Scientist, Automotive
Lender Decision ML Model for E-Commerce Website

- Review past 2 years of lender credit information data including user's income, credit score (range), rate of interest, approval status, etc.
- Create statistical analysis of raw data and pick possible features.
- Test possible features against AWS SageMaker autopilot with a target variable
- Test possible features against regression algorithm like XGBoost etc.

- Deploy model with AWS SageMaker for realtime inference.

EDUCATION

Lamar University

Master's Degree, Computer Science (2015 - 2017)

New England College

Master of Science - MS, Data Science (June 2020 - March 2022)

Gujarat Technological University, Ahmedbabd

Bachelor's Degree, Computer Engineering (2010 - 2014)

RESOURCE 3: DATA ENGINEER

RESOURCE SUMMARY

Seasoned data engineer, specializing in leveraging the power of public cloud services, particularly AWS and Azure, to design and implement cutting-edge data platforms. Expertise in advanced data technologies such as Databricks and Snowflake, enabling organizations to harness the full potential of their data.

AREAS OF EXPERTISE

- Apache Parquet
- Apache Cassandra
- Apache Kafka
- Apache Flink
- AWS Rekognition
- MS Power BI
- Apache Hudi
- AWS EMR
- CPython
- PySpark
- AWS Step Functions
- Snowflake
- AWS DynamoDB
- AWS Lake Formation
- AWS Athena
- AWS Lambda
- AHashiCorp Terraform
- AWS Key Management Service (KMS)
- AWS Database Migration Service
- Apache Sqoop
- Apache Hive
- Apache Oozie
- Apache Hadoop
- Apache Spark
- Databrick Lakehouse Platform
- AWS Kinesis
- AWS SageMaker
- AWS AppFlow
- Tableau
- Python
- AWS RDS
- AWS Redshift
- AWS Glue
- AWS Data Pipeline
- AWS CloudFormation
- AWS Storage Gateway
- AWS EventBridge
- AWS Aurora
- AWS DataSync

SKILLS OVERVIEW

- Strong business and communication skills as well as knowledge of the big data industry
- Machine learning, predictive modeling and natural language processing
- Knowledge of data governance and security best practices
- Data visualization, data migration and data modeling
- Preparing design specifications, architecture blueprints, frameworks, data flow diagrams, and decision records that support project teams and client stakeholders.

RELEVANT EXPERIENCE

10/2024-03/2025

Data Architect, Oracle to Snowflake Migration for MassHealth

- Developed source data extraction pipelines using Informatica to automate data extraction, and loading into Snowflake, ensuring data integrity and minimizing downtime.
- Optimized SQL queries and schema design for Snowflake's columnar storage and micro-partitioning, significantly improving query performance and storage efficiency.
- Collaborated with cross-functional teams, including DBAs, data engineers, and business analysts, to ensure a seamless transition with minimal impact on reporting and business operations.
- Designed and executed validation frameworks to perform data reconciliation between Oracle and Snowflake, ensuring data accuracy, completeness, and consistency post code migration.

06/2024-10/2024

Data Architect, Unified Data Share for Hillsborough County Clerk

- Led the design and development of enhanced Data Share integration services leveraging Azure GOV Cloud and Boomi's automation platform, ensuring secure and scalable data exchange.
- Expanded Data Share file services by improving metadata tracking and search capabilities, enabling efficient data discovery and governance.
- Enhanced SQL Database utilization by incorporating message share capabilities to capture and manage metadata and variable messaging services, improving data consistency and accessibility.
- Worked closely with stakeholders to prioritize and refine future state feature requirements, translating them into actionable development tasks within JIRA.
- Provided technical guidance and support to the Hillsborough team as they onboarded and adopted the current state UDS platform, ensuring seamless integration and alignment with business needs.

03/2023-05/2024

Data Architect, Flour-Milling and Ingredient Company

Insight and Modern Data Platform

- Developing the architectural blueprint for the modern data platform, ensuring scalability, performance, and security.
- Overseeing the implementation of the data platform on Azure and Databricks, including setting up infrastructure, configuring services
- Designing and implementing data models that support the company's analytical and operational needs, ensuring data integrity and efficiency
- Designing and implementing data ingestion pipelines to efficiently and reliably ingest data from various sources into the data platform
- Establishing data governance policies and processes to ensure data quality

- Leading a team of data engineers and analysts, providing guidance, mentoring, and technical expertise to ensure the successful implementation
- Collaborating with stakeholders across the organization to understand their data requirements and ensure that the data platform meets their needs
- Evaluating new technologies and tools that can enhance the capabilities of the data platform and improve overall efficiency
- Identifying opportunities for continuous improvement in the data platform architecture, processes, and technologies to drive innovation and efficiency

08/2022-06/2023

Data Architect, Financial and Insurance

Rehost On-Prem SQL Server Cluster to AWS

- Develop migration plan, assess risks, ensure smooth transition to cloud with ZERO downtime.
- Collaborate for setting up infrastructure on cloud, ensure it meets requirements.
- Design strategy for ZERO downtime, select tools for data integrity and efficiency.
- Create detailed documentation of migration plan, configuration, and validation.
- Provide support post-migration, address issues for smooth transition.

01/2021-06/2022

Data Architect, Automotive

Enterprise Data Platform

- Plan the development of enterprise data platform on AWS, considering scalability and security.
- Collaborate with the infrastructure team to set up scalable and secure infrastructure.
- Design and implement efficient data ingestion pipelines for SAP, Salesforce, and OLTP
- Develop data processing workflows using AWS Glue and Athena for analytics.
- Design storage solutions using S3 and RDS for structured and unstructured data.
- Establish data governance policies using AWS Lake Formation and Glue for cataloging data.
- Lead a team of data engineers and developers, providing guidance and support for project success.

04/2020-01/2021

Senior Data Scientist, Automotive

Lender Decision ML model for E-Commerce Website

- Review past 2 years of lender credit information data including user's income, credit score(range), rate of interest, approval status etc.
- Create statistical analysis of raw data and pick possible features.
- Test possible features against AWS SageMaker autopilot with a target variable
- Test possible features against regression algorithm like XGBoost etc.
- Deploy model with AWS SageMaker for realtime inference.

5 COSTS

Provide the total not-to-exceed costs for providing the Services based on projected hours, proposed hourly rates, as well as any other appropriate costs, by program phase in the Budget Template (Attachment C). List additional fees, overhead charges, or reimbursable expenses, if any. As a general policy, the Mass Tech Collaborative does not pay markups on reimbursables or out-of-pocket expenses. The Mass Tech Collaborative also does not pay for word processing, overtime or meals. For travel costs, the Mass Tech Collaborative pays the IRS rate per mile.

Response:

The total not-to-exceed cost for providing these services, based on projected hours, proposed hourly rates, and other appropriate costs, is outlined by program phase in the completed Budget Template (Attachment C). This cost proposal reflects a realistic and cost-effective approach to delivering the full scope of work.

No additional fees, overhead charges, or markups on reimbursable or out-of-pocket expenses have been included, in accordance with MassTech's policies. Furthermore, no costs have been allocated for word processing, staff overtime, or meals. Travel costs, when applicable, have been calculated at the current IRS mileage reimbursement rate.

This budget is designed to maximize cost-effectiveness while maintaining the quality and timeliness of deliverables.

6 PAST PERFORMANCE

Three examples of work previously performed by the Respondent that are substantially similar to the Services. Respondent shall also provide references including a contact person, email and phone number

Response:

The following Presidio clients are happy to serve as references but prefer not to receive unscheduled inquiries. As a courtesy, we request that MassTech coordinate any contact with the individuals listed below via Presidio's MassTech Account Manager, Dan Buckley.

COMMONWEALTH OF MASSACHUSETTS DEPARTMENT OF MENTAL HEALTH (DMH)	
ADDRESS	25 Staniford Street, Boston, MA 02114
CONTACT NAME & TITLE	Brigido Ramirez, Assistant Commissioner for Data and Analytics
PHONE NUMBER	617-935-6080
EMAIL	Brigido.Ramirez2@mass.gov
SIZE / SCOPE OF PROJECT	Presidio and DMH have been working together for the past year to develop and deploy and migrate into a modern data platform and working iteratively to support key use case identification and rollout. Presidio has partnered with DMH and EHS to prioritize both use cases and critical data sources, including the EHR system. We have provided migration services around data architecture, data engineering, AWS DevOps expertise, data governance, BI design and development, and overall project/program management.

COMMONWEALTH OF MASSACHUSETTS - MASSHEALTH	
ADDRESS	100 Hancock Street, 1st Floor, Quincy, MA 02171
CONTACT NAME & TITLE	Keith Raboin, CDIO EOHHS
PHONE NUMBER	617-855-5309
EMAIL	keith.b.raboin@mass.gov
SIZE / SCOPE OF PROJECT	Presidio has been working with the Commonwealth of MA to migrate and modernize the MassHealth Enterprise Data Warehouse (EDW) which was successfully completed in June 2025. We have partnered tightly with the HHS/MassHealth team since 2022 and continue with additional activities well into next year and beyond. Our team has been seen as an extension to the Commonwealth's capabilities as we co-deliver with fully integrated teams to help the Commonwealth migrate and modernize across people, process, and technology. Our migration project team worked with the technology solutions of AWS, Snowflake, Informatica, and Tableau throughout the course of this engagement.

COMMONWEALTH OF MASSACHUSETTS – DEPARTMENT OF PUBLIC HEALTH (DPH)	
ADDRESS	250 Washington Street, Boston, MA 02108
CONTACT NAME & TITLE	Maura Brown, EOHHS Director of Data Strategy Rebecca White, DMI Portfolio and Project Manager
PHONE NUMBER	Maura Brown: 857-260-1108 Rebecca White: 857-408-2030
EMAIL	Maura.B.Brown@mass.gov Rebecca.H.White@mass.gov
SIZE / SCOPE OF PROJECT	The Presidio team was engaged to do knowledge transfer and takeover/ complete the critical Public Health Data warehouse (PHD) migration project from another integrator and has been partnering with the DPH and EHS teams to progress the initiative since February 2025. We have been working with an integrated EHS/Presidio engineering team to plan and execute each part of the PHD migration and are scheduled to wrap up later this year. Key technologies in use during the migration were AWS and Snowflake, and our teams are always aligning with HHS and TSS leading practices.

7 W-9

Respondent's W-9.

Response:

Presidio includes a copy of our W-9 on the pages that follow.

PRESIDIO™

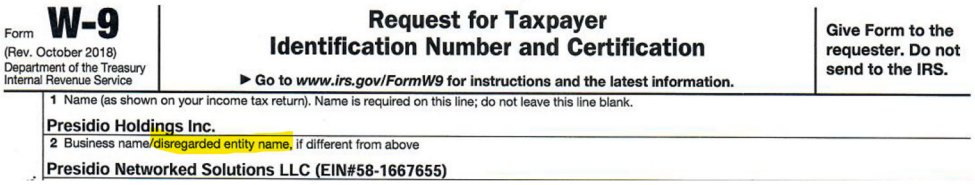
Presidio Networked Solutions LLC
W-9 Attachment

Business Name:	**Presidio Networked Solutions LLC**
Federal Tax classification:	**SMLLC Disregarded Entity**
Tax Address*:	**12100 Sunset Hills Road, Suite 300, Reston VA 20190**
Taxpayer Identification Number (TIN):	**58-1667655**

Presidio Networked Solutions LLC ("PNS") should be the vendor on file and the tax identification number **58-1667655** for the purpose of receiving **customer payments**. This is the operating legal entity that is in the contracts.

Presidio Networked Solutions LLC is a single-member LLC (SMLLC) that is treated as a disregarded entity for U.S. income tax purposes.

Per IRS instructions, the disregarded entity MUST be shown on **line 2** of Federal Form W-9. See Instructions on IRS website. Under penalties of perjury, the form must be filled out as instructed by the IRS.



Form **W-9** (Rev. October 2018) Department of the Treasury Internal Revenue Service

Request for Taxpayer Identification Number and Certification

▶ Go to *www.irs.gov/FormW9* for instructions and the latest information.

Give Form to the requester. Do not send to the IRS.

1 Name (as shown on your income tax return). Name is required on this line; do not leave this line blank.

Presidio Holdings Inc.

2 Business name/disregarded entity name, if different from above

Presidio Networked Solutions LLC (EIN#58-1667655)

Presidio Holdings Inc. EIN 45-1263428 is the single owner, and as such, is ***required*** to be shown on line 1 of IRS Form W-9 or similar State Information Reporting Forms. The owner's EIN 45-1263428 is to be used for federal tax reporting purposes and not the disregarded entity. See IRS Regulations section 301.7701-2(c)(2)(iii)).

***Per IRS instructions**, the address on any TAX forms is the tax headquarters listed above.

Please remit to "Presidio Networked Solutions LLC" as the "Payee":

For Payments by Mail: Presidio Networked Solutions LLC PO Box 822169 Philadelphia, PA 19182-2169	**For Payments by Overnight Courier:** PNC Bank/Presidio Networked Solutions LLC Attn: 822169 525 Fellowship Rd, Suite 330 Mt Laurel, NJ 08054-3415
For Payments via Wire: Receiving Bank: PNC Bank PNC Bank ABA: 031000053 Beneficiary: Presidio Networked Solutions LLC Beneficiary Account Number: 8611678714	

If you have any questions about the bank account for payment, please contact DLAccountsReceivable@Presidio.com. Should you have any questions about your invoice, see your Collections Specialist's name, phone and fax number located on your invoice.

Form **W-9**
(Rev. March 2024)
Department of the Treasury
Internal Revenue Service

Request for Taxpayer Identification Number and Certification

Go to *www.irs.gov/FormW9* for instructions and the latest information.

Give form to the requester. Do not send to the IRS.

Before you begin. For guidance related to the purpose of Form W-9, see *Purpose of Form*, below.

Print or type. See *Specific Instructions* on page 3.

1 Name of entity/individual. An entry is required. (For a sole proprietor or disregarded entity, enter the owner's name on line 1, and enter the business/disregarded entity's name on line 2.)

Presidio Holdings Inc.

2 Business name/disregarded entity name, if different from above.

Presidio Networked Solutions LLC (EIN#58-1667655)

3a Check the appropriate box for federal tax classification of the entity/individual whose name is entered on line 1. Check only **one** of the following seven boxes.

☐ Individual/sole proprietor ☑ C corporation ☐ S corporation ☐ Partnership ☐ Trust/estate

☐ LLC. Enter the tax classification (C = C corporation, S = S corporation, P = Partnership) . . . . .

Note: Check the "LLC" box above and, in the entry space, enter the appropriate code (C, S, or P) for the tax classification of the LLC, unless it is a disregarded entity. A disregarded entity should instead check the appropriate box for the tax classification of its owner.

☐ Other (see instructions)

4 Exemptions (codes apply only to certain entities, not individuals; see instructions on page 3):

Exempt payee code (if any)

Exemption from Foreign Account Tax Compliance Act (FATCA) reporting code (if any)

(Applies to accounts maintained outside the United States.)

3b If on line 3a you checked "Partnership" or "Trust/estate," or checked "LLC" and entered "P" as its tax classification, and you are providing this form to a partnership, trust, or estate in which you have an ownership interest, check this box if you have any foreign partners, owners, or beneficiaries. See instructions . . . . . ☐

5 Address (number, street, and apt. or suite no.). See instructions.

12100 Sunset Hills Road, Suite 300

Requester's name and address (optional)

6 City, state, and ZIP code

Reston, VA 20190

7 List account number(s) here (optional)

Part I Taxpayer Identification Number (TIN)

Enter your TIN in the appropriate box. The TIN provided must match the name given on line 1 to avoid backup withholding. For individuals, this is generally your social security number (SSN). However, for a resident alien, sole proprietor, or disregarded entity, see the instructions for Part I, later. For other entities, it is your employer identification number (EIN). If you do not have a number, see *How to get a TIN*, later.

Note: If the account is in more than one name, see the instructions for line 1. See also *What Name and Number To Give the Requester* for guidelines on whose number to enter.

Social security number

			–			–				

or

Employer identification number

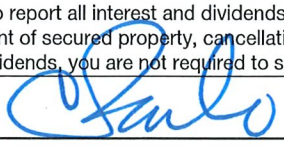
4	5	–	1	2	6	3	4	2	8

Part II Certification

Under penalties of perjury, I certify that:

1. The number shown on this form is my correct taxpayer identification number (or I am waiting for a number to be issued to me); and
2. I am not subject to backup withholding because (a) I am exempt from backup withholding, or (b) I have not been notified by the Internal Revenue Service (IRS) that I am subject to backup withholding as a result of a failure to report all interest or dividends, or (c) the IRS has notified me that I am no longer subject to backup withholding; and
3. I am a U.S. citizen or other U.S. person (defined below); and
4. The FATCA code(s) entered on this form (if any) indicating that I am exempt from FATCA reporting is correct.

Certification instructions. You must cross out item 2 above if you have been notified by the IRS that you are currently subject to backup withholding because you have failed to report all interest and dividends on your tax return. For real estate transactions, item 2 does not apply. For mortgage interest paid, acquisition or abandonment of secured property, cancellation of debt, contributions to an individual retirement arrangement (IRA), and, generally, payments other than interest and dividends, you are not required to sign the certification, but you must provide your correct TIN. See the instructions for Part II, later.

Sign Here | **Signature of U.S. person**  **Date** 1/1/2025

General Instructions

Section references are to the Internal Revenue Code unless otherwise noted.

Future developments. For the latest information about developments related to Form W-9 and its instructions, such as legislation enacted after they were published, go to *www.irs.gov/FormW9*.

What's New

Line 3a has been modified to clarify how a disregarded entity completes this line. An LLC that is a disregarded entity should check the appropriate box for the tax classification of its owner. Otherwise, it should check the "LLC" box and enter its appropriate tax classification.

New line 3b has been added to this form. A flow-through entity is required to complete this line to indicate that it has direct or indirect foreign partners, owners, or beneficiaries when it provides the Form W-9 to another flow-through entity in which it has an ownership interest. This change is intended to provide a flow-through entity with information regarding the status of its indirect foreign partners, owners, or beneficiaries, so that it can satisfy any applicable reporting requirements. For example, a partnership that has any indirect foreign partners may be required to complete Schedules K-2 and K-3. See the Partnership Instructions for Schedules K-2 and K-3 (Form 1065).

Purpose of Form

An individual or entity (Form W-9 requester) who is required to file an information return with the IRS is giving you this form because they

8 AUTHORIZED APPLICATION SIGNATURE ACCEPTANCE FORM

Authorized Application Signature and Acceptance Form (Attachment B). By executing the Authorized Respondent's Signature and Acceptance Form and submitting a response to this RFP, Respondents certify that they (1) are in compliance with the terms, conditions and specifications contained in this RFP, (2) acknowledge and understand the procedures for handling materials submitted to the Mass Tech Collaborative as set forth in subsection d. below, (3) agree to be bound by those procedures, and (4) agree that the Mass Tech Collaborative shall not be liable under any circumstances for the disclosure of any materials submitted to the Mass Tech Collaborative pursuant to this RFP or upon the Respondent's selection.

Response:

Presidio acknowledges and agrees to the terms outlined in the Authorized Application Signature and Acceptance Form (Attachment B). By submitting the completed and signed Attachment B, Presidio certifies its compliance with the RFP's terms, conditions, and specifications; acknowledges and accepts the procedures for handling submitted materials as described in subsection d; and agrees that Mass Tech Collaborative shall not be held liable for any disclosure of materials submitted pursuant to this RFP or upon Presidio's selection.

The signed Attachment B follows.

Attachment B

Massachusetts Technology Collaborative
Authorized Respondent's Signature and Acceptance Form

The undersigned is a duly authorized representative of the Respondent listed below. The Respondent has read and understands the RFP requirements. The Respondent acknowledges that all of the terms and conditions of the RFP are mandatory, and that Respondent's response is compliant with such requirements.

The Respondent understands that, if selected by the Mass Tech Collaborative, the Respondent and the Mass Tech Collaborative will execute an Agreement specifying the mutual requirements of participation.

Respondent agrees that the entire bid response will remain valid for sixty (60) days from receipt by the Mass Tech Collaborative.

I certify that Respondent is in compliance with all corporate filing requirements and State tax laws.

I certify that the statements made in this response to the RFP, including all attachments and exhibits, are true and correct to the best of my knowledge.

I acknowledge and understand the procedures for handling confidential materials submitted to the Mass Tech Collaborative as set forth in Section 3.1(d); I agree to be bound by those procedures and agree that the Mass Tech Collaborative shall not be liable under any circumstances for the disclosure of any materials submitted to the Mass Tech Collaborative pursuant to this RFP or upon the Respondent's selection.

Respondent: Presidio Networked Solutions LLC
(Printed Name of Respondent)

By:
(Signature of Authorized Representative)

Name: James J. Doyle
Title: VP, Support Services
Date: 9/30/25

Massachusetts Technology Collaborative

Grant Agreement
Standard Budget and Invoice Template
Invoiced Costs

Line Number	Budgeting																	
2	Applicant Information																	
3	Applicant: Presidio Networked Solutions LLC			Procurement No.: RFP No. 2026-JAII-01														
4	Budget Period:			Title of Proposed Project:														
5	Address: 10 Sixth Road, Woburn, MA 01801			AI Hub Data Commons Collaborative Consulting Services														
6																		
7																		
8				MTC Funding Re	1,447,800													
9				Total Project Cost	1,447,800													
10	Cost Elements				Total Project	Invoice 1		Invoice 2		Invoice 3		Invoice 4		Cumulative				
11						Period: MM/DD/YYYY-MM/DD/YYYY		Period: MM/DD/YYYY-MM/DD/YYYY		Period: MM/DD/YYYY-MM/DD/YYYY		Period: MM/DD/YYYY-MM/DD/YYYY		Period: MM/DD/YYYY-MM/DD/YYYY				
12							Amount		Amount		Amount		Amount	Expended		Remaining		
13	I. Direct Labor																	
14	*name/title*	*HRS*	*Rate*			*HRS*		*HRS*		*HRS*		*HRS*						
15	Solution Owner	400	$275	$110,000		$ -		$ -		$ -		$ -		$ -		$	110,000.00	
16	Data Architect	640	$330	$211,200		$ -		$ -		$ -		$ -		$ -		$	211,200.00	
17	Senior Engineer (x4)	2560	$285	$729,600		$ -		$ -		$ -		$ -		$ -		$	729,600.00	
18	Business Analyst	280	$225	$63,000		$ -		$ -		$ -		$ -		$ -		$	63,000.00	
19	QA Engineer	440	$250	$110,000		$ -		$ -		$ -		$ -		$ -		$	110,000.00	
20						$ -		$ -		$ -		$ -		$ -		$	-	
21	Total Direct Labor			$1,223,800	0	$ -	0	$ -	0	$ -	0	$ -		$ -		$	1,223,800.00	
22																		
23	II. Direct Labor Fringe Cost	Rate (%):		$0		$ -		$ -						$ -		$	-	
24																		
25	II. Subcontractors/Consultants/Contracted Staffing																	
26	*Firm name/title or role/type of services to be provided*	*HRS*	*Rate*			*HRS*		*HRS*		*HRS*		*HRS*						
27				$0		$ -		$ -		$ -		$ -		$ -		$	-	
28				$0		$ -		$ -		$ -		$ -		$ -		$	-	
29				$0		$ -		$ -		$ -		$ -		$ -		$	-	
30				$0		$ -		$ -		$ -		$ -		$ -		$	-	
31				$0	5	$ -		$ -		$ -		$ -		$ -		$	-	
32				$0		$ -		$ -		$ -		$ -		$ -		$	-	
33	Total Subcontractors/Consultants			$0	5	$ -	0	$ -	0	$ -	0	$ -		$ -		$	-	
34																		
35	III. Direct Materials (provide description and costs basis)																	
36	Curriculum Development & Training			$0										$0		$	0	
37				$0										$0		$	0	
38				$0										$0		$	0	
39				$0										$0		$	0	
40				$0										$0		$	0	
41				$0										$0		$	0	
42				$0										$0		$	0	
43	Total Direct Materials			$0			$0		$0		$0		$0	$0		$	0	
44																		
45	IV. Travel													$0		$	0	
46																		
47	V. Other Direct Costs *(list type and description such as venue rental costs and cost basis)*																	
48	Snowflake Data Platform Operations Cost (Annual)			$200,000										$ -		$	200,000.00	
49	AWS Cloud Operations Cost (Annual)			$24,000										$ -		$	24,000.00	
50				$0										$ -		$	-	
51				$0										$ -		$	-	
52				$0										$ -		$	-	
53	Total Other Direct Costs			$224,000		$ -		$ -		$ -		$ -		$ -		$	224,000.00	
54																		
55	VI. General & Administrative Expense/Overhead	Rate (%):		$0		$ -		$0		$0		$0		$ -			$0	
56																		
57	**Total PROJECT COSTS**			**$1,447,800**		**$0**		**$0**		**$0**		**$0**		**$0**		**$**	**1,447,800**	

LEGEND:

Formula cell
Data entry cell

This section is for post-contract invoicing only. Do not complete with proposal.